ANALYSIS OF THE SEQUENCES REQUIRED FOR TRANSCRIPTIONAL REGULATION OF A HUMAN H4 HISTONE GENE IN VIVO

By

PAUL EDMOND KROEGER

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1988

ACKNOWLEDGEMENTS

I would like to thank Janet and Gary Stein for the opportunity to work in their laboratory and explore molecular biology from a great many perspectives. I also appreciate the advice and encouragement of my other committee members, Drs. Ostrer, Hauswirth and Moyer.

The Stein's laboratory has been filled with many characters over these last six years and I owe thanks to all of them. I would like to thank Farhad Marashi, Mark Plumb, and Linda Green (especially Linda) for their technical expertise and friendship. My fellow graduate students Gerard Zambetti, Dave Collart, André van Wijnen, and Anna Ramsey, I thank for their comradeship during the preceding years. The laboratory would not have been the same without Charles Stewart, Urs Pauli, and Sue Chrysogelos all of whom have given me new perspectives on life and science. I thank Tim Morris for his constant good nature, advice, and stimulating conversions (although we did not always agree). I would particularly like to thank Ken Wright for our many successful collaborative adventures in the laboratory, his friendship, and generosity when it was most needed.

Finally I thank my wife Carol, and our new son, Alan, who have given me constant inspiration to continue down what has been a long and unusual path through graduate school. My parents have also been a constant source of advice and encouragement, and I thank them for their unending interest.

TABLE OF CONTENTS

KEY TO ABBREVIATIONS

ATP:	Adenosine 5'-triphosphate
bp:	Base pair
C:	Centigrade
CIP:	Calf intestinal phosphatase
CTP:	Cytidine 5'-triphosphate
DEPC:	Diethylpyrocarbonate
DNA:	Deoxyribonucleic acid
DNase I:	Deoxyribonuclease I
DU:	Densitometry units
EDTA:	Disodium Ethylenediaminetetraacetate
EGTA:	Ethylenebis(oxyethylenenitrilo)tetraacetic acid
GTP:	Guanosine 5'-triphosphate
Hepes:	N-2-hydroxyethylpiperizine-N'-2-ethanesulfonic acid
HU:	Hydroxyurea
l:	Liter
M:	Molar
μCi:	Microcurie
mg:	Milligram
μg:	Microgram
ml:	Milliliter
μl:	Microliter

mM:	Millimolar
mRNA:	Messenger ribonucleic acid
nm:	Nanometer
nt:	Nucleotides
OD:	Optical density
Pipes:	[1,4-piperazinebis(ethanesulfonic acid)]
PVS:	Polyvinylsulfate
RNA:	Ribonucleic acid
RNaseA:	Ribonuclease A
rpm:	Revolutions per minute
SDS:	Sodium dodecyl sulfate
SV40	Simian virus 40
TCA:	Trichloroacetic acid
Tris:	Tris(hydroxymethyl)aminomethane + Hydrochloric acid
TTP:	Thymidine 5'-triphosphate

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

ANALYSIS OF THE SEQUENCES REQUIRED FOR TRANSCRIPTIONAL REGULATION OF A HUMAN H4 HISTONE GENE IN VIVO

By

Paul Edmond Kroeger

August 1988

Chairman: Janet Stein
Major Department: Immunology and Medical Microbiology

We have characterized the sequences required for the transcriptional regulation of the FO108 human H4 histone gene in vivo. Recombinant cell lines that contained deletion constructs of the H4 promoter region were prepared in mouse C127 cells, and the level of human H4 histone gene expression was measured by S1 nuclease analysis. We found that the minimal sequences required for the initiation of transcription from this gene were contained within the 73 nucleotides 5′ to the initiation site of transcription. Within this region are located an in vivo protein binding site (Site II), the GGTCC element and the TATA box. Deletion of the distal half of Site II abolished site specific initiation of transcription and demonstrated that the TATA box and GGTCC element were not sufficient for initiation in vivo. Extension of the H4 promoter to -100 base pairs resulted in a significant increase in transcription and this increase correlated with the

presence of an Sp1 site in the proximal half of the upstream protein binding site, Site I. If the promoter region was lengthened to -410 nucleotides, there was a two-fold increase in the level of transcription. Deletion analysis suggested that the "distal-proximal" positive element was located from in the region from -210 to -330 base pairs 5' to the cap site. We investigated the functionality of a previously identified enhancer-like element located very far upstream in the pFO116 fragment of λ HHG 41 and demonstrated that although it functioned in HeLa cells it was not functional in mouse C127 cell lines.

S1 analysis of distal deletion constructs supported the idea that a negative regulatory element of H4 gene transcription was located between nucleotides -730 and -1010. Analysis of the region demonstrated consensus sequences for a topoisomerase II site, nuclear matrix attachment sites, and a very high A/T content (70%) suggestive of bent DNA. Taken together this set of results implied that the DNA topology of this region might be important for H4 gene regulation.

Additional studies demonstrated that Alu repetitive sequences in the histone deletion constructs could mediate specific integration into the mouse chromosome and that high copy number was possible.

CHAPTER I

INTRODUCTION

The goal of this study has been to assess the contribution of promoter sequences in the FO108 human H4 histone gene 5' flanking region to transcriptional regulation of the gene. We have endeavored to define the sequences necessary for the initiation and augmentation of transcription. The TATA box, GGTCC element, "CAAT box," "CCAAT box," and Sp1 site have been implicated in transcriptional regulation and are reviewed below. We have also investigated a putative enhancer-like element and negative regulatory sequence and so these sequences are also discussed below.

Historical Background

The concepts governing gene regulation, as we know them today, have their foundations in the work of many biochemists and geneticists who introduced the ideas of positive and negative regulation in prokaryotic gene expression. The observations of many, that the total genetic potential of a cell was never expressed simultaneously, referred to as "genetic adaptation," led Jacob and Monod (1961) to address the question of what controls this phenomenon. In their seminal paper the "operon model" was proposed. This model described how structural genes expressed themselves and how that expression was regulated. It had been known for some time that bacteria could respond to various

nutrients by synthesizing new metabolic enzymes, so Jacob and Monod investigated the lactose metabolic pathway of Escherichia coli (E. coli). Their work was encouraged by many earlier investigators, including Demerac (1956), who made the observation that genes coding for similar enzymatic function were located in localized regions of the Salmonella chromosome. Demerac was able to conclude that the genes he had investigated were in a nonrandom distribution and that perhaps this conferred an evolutionary advantage to the organism.

The lac operon is one of the most well studied genetic systems in all of prokaryotic and eukaryotic molecular biology. The many intuitive observations and predictions of Jacob and Monod and colleagues led to the identification of the components of the lac operon: the repressor, produced by the lac I gene; the lac operator, promoter, and three linked structural genes. The interplay of inducer and repressor was demonstrated, and Jacob and Monod proposed that the lac operon was subject to negative regulation. An initial observation of Jacob and Monod (1961) was that the control gene would make repressors that would turn off the structural genes. The isolation of nonsense mutations in the lac I gene (Bourgeois et al., 1965) provided convincing evidence for the nature of repressors. Suppression of the nonsense mutation restored repressor function and demonstrated that repressor genes encoded repressor proteins. The final proof was the isolation of the lac repressor by Gilbert and Müller-Hill (1966). In addition it was demonstrated that the lac operon and others were under more general control by catabolite activator protein and 3'5'-cyclic-

AMP as it was shown that both are required, in addition to the inducing molecule, for the operon to be transcribed (Emmer et al., 1970).

The ensuing years have led to refinement of the operon model as well as its acceptance as one of the general organizational patterns characteristic of prokaryotes. In particular, the concepts of protein/DNA interactions, repression, and positive and negative regulation have carried over into eukaryotic molecular biology and have served as a basis for unraveling the complexity of the eukaryotic cell. The extension of these ideas has allowed considerable progress; however, the original view that all genes, prokaryotic and eukaryotic, would have similar regulatory and organizational patterns has not been borne out. In fact there is a great diversity in the regulatory mechanisms that govern both prokaryotic and eukaryotic gene expression.

The control of eukaryotic gene regulation has been of obvious interest, but research has been slower than in prokaryotes because of the complexity and technical difficulties encountered when working with the eukaryotic cell. Two avenues of study have predominated in eukaryotic molecular biology: the investigation of viral models such as adenovirus and SV40 (as was done with the prokaryotic phages lambda and T7) and the characterization of cellular genes and the proteins that regulate their expression.

Eukaryotic molecular biologists have had to develop the appropriate technology because many of the advantageous prokaryotic techniques are not directly applicable to eukaryotic systems. Two of the most important discoveries that have revolutionized molecular biology are

restriction enzymes (reviewed by Nathans and Smith, 1975) and DNA ligase (Modrich et al., 1973; Weiss and Richardson, 1967). With these new enzymatic tools the ability to manipulate DNA fragments developed quickly and was responsible for the present state of advancement.

Viral Model Systems

The utilization of viral model systems for the characterization of eukaryotic regulatory mechanisms was a logical extension of the work done in prokaryotes. In particular, adenovirus and SV40 have provided considerable insights into eukaryotic gene regulation. Without an understanding of the exact mechanisms involved in the various processes of RNA transcription and DNA replication, it was obvious to early investigators that viruses, such as SV40, could invade and eventually kill the host cell and yet were extremely dependent on the cell's enzymatic machinery to accomplish their replicative cycle.

Adenoviruses were first isolated by Rowe et al. (1953) as the agent responsible for the degeneration of human adenoid tissue in culture. The adenovirus life cycle in human cells has been examined with respect to the virus-specific proteins produced, replication of viral DNA, transcription of viral genes, and effect on the host cell (Reviewed in Tooze, 1980). Initial studies demonstrated that there were two phases--early and late--in the expression of adenovirus genes (Lindberg et al., 1972). As a measure of the impact of infection on the cell, adenovirus mRNA comprises almost all the mRNA bound to polyribosomes by the end of the replicative cycle (Thomas and Green, 1966). The early viral mRNA was detected and mapped to precise locations on the adenovirus genome by R-loop mapping (Thomas et al.,

1976) and hybridization to restriction endonuclease fragments of adenovirus DNA (Sharp et al., 1975). Restriction enzymes permitted the mapping and orientation of DNA fragments and transcription units on the SV40 genome as well (Khoury et al., 1973; Sambrook et al., 1973).

Several laboratories utilized adenovirus/SV40 recombinant hybrids to define essential genomic regions of each. In particular, the hybrid viruses were useful in the determination of the functional "helper" domain of the SV40 T antigen, as adenovirus requires "help" to grow in nonpermissive cells (Fey et al., 1979). With the mRNA coding regions mapped on the adenovirus and SV40 genomes, a more informative analysis and interpretation were initiated which have begun to elucidate the complex nature of transcriptional regulation in these viruses. The promoter structure and presence of enhancing/silencing elements in these viruses have served as continuing models for studies of cellular promoters and regulatory sequences. Additionally, although not discussed here, both adenovirus and SV40 were utilized in the discovery of mRNA splicing (Berk and Sharp, 1977, 1978), which has revolutionized our concepts of gene regulation and expression.

Chromatin Studies

At the same time that the viral model systems were beginning to be reasonably well understood, there were a number of investigators pursuing the characterization of cellular genes and transcriptional mechanisms. Although restriction enzymes had been discovered (Smith and Wilcox, 1970) and their applicability realized, it was several years before their purification and recombinant DNA technology were worked out to make them sufficiently useful. This lag did not deter a

number of investigators from direct examination of the transcriptional process in eukaryotic cells. As early as 1962 isolated pea embryo chromatin had been utilized as a template for transcription (Huang and Bonner, 1962). Isolated chromatin was incubated with bacterial RNA polymerase (the purification of eukaryotic RNA polymerases had not been achieved at this time) and the four ribonucleoside triphosphates. A comparative analysis of transcription from chromatin and deproteinized DNA of the same source indicated that the chromatin was less able to support transcription (Huang and Bonner, 1962). It was postulated that part of the chromatin was repressed, perhaps due to the presence of histone proteins bound to the DNA. The amount of transcription possible from a known quantity of chromatin was referred to as its template capacity. The determination of template capacity in chick oviduct, a steroid responsive tissue, led to the observation that the level of transcription was modulated with the addition of hormone (Dahmus and Bonner, 1965). The amount of template capacity also correlated with the various developmental stages of sea urchin growth (Johnson and Hnilica, 1970). Another more accurate measure of the "transcriptional capacity" of a sample of chromatin was the number of RNA polymerase initiation sites. Cedar and Felsenfeld (1973) first measured the number of E. coli RNA polymerase initiation sites on chromatin by incubating chromatin and RNA polymerase together in the absence of ribonucleoside triphosphates. Next, the addition of the ribonucleoside triphosphates with high levels of ammonium sulfate permitted elongation but not reinitiation. One of the major criticisms of this early work was that the use of bacterial RNA polymerase made an

accurate interpretation in doubt. Comparative studies were performed by Mandel and Chambon (1970) and Tsai et al. (1976). These investigators demonstrated that there was no competition for either SV40 DNA or calf thymus DNA by the bacterial or eukaryotic RNA polymerase. However, when Tsai et al. (1976) compared hen oviduct and E. coli RNA polymerase initiation sites on chick DNA or chick oviduct chromatin, they found no competition on the DNA, but direct competition in the chromatin sample. Thus it appeared that chromosomal proteins could modify the initiation specificity such that both enzymes were competing for similar sites. To establish this point conclusively, the product mRNAs had to be examined. Filter hybridization techniques were developed that permitted the detection of reiterated gene transcripts and particularly abundant mRNAs. At the level of sensitivity possible with this methodology, in vitro chromatin transcription appeared to reflect an accurate view of the transcriptional status in vivo (Bacheler and Smith, 1976).

The next major advance was the fractionation of chromosomal proteins in an effort to reconstitute transcriptionally competent DNA into chromatin in vitro. The first attempts to reconstitute chromatin were studies by Paul and Gilmour (1966, 1968) and Bekhor et al. (1969) in which they fractionated chromatin proteins in an attempt to discover what group of proteins controlled transcriptional. Their results indicated that the non-histone chromosomal protein (NHCP) fraction was probably responsible. The role of NHCP in the expression

of several genes has been reviewed (Stein et al., 1974; Simpson, 1973).

Experiments became more refined as exemplified by the studies of Tsai et al. (1976) who examined the inducible ovalbumin gene in the chick oviduct system. The role of NHCP was established, and through a series of competition assays with induced and uninduced NHCPs it was demonstrated that in vitro expression of the ovalbumin gene was stimulated by the appearance, upon steroid induction, of a positive regulatory factor. Histones, a moderately reiterated family of genes (Stein et al., 1984), were also studied in a similar manner to examine the role of NHCPs. Several studies indicated that NHCPs were involved in the increased expression of the histone genes during S-phase of the cell cycle (Park et al., 1976; Stein et al., 1975). Kleinsmith et al. (1976) extended the characterization and demonstrated that phosphorylation of the NHCP was necessary for optimal in vitro expression of the histone genes. When the NHCPs were treated with phosphatase before addition to the reaction, there was a decrease in the number of transcription initiation sites.

The role of the histone proteins in transcription has been of great interest because they form such a close association with the DNA. Studies with either electron microscopy or nuclease digestion have demonstrated that there is either a change in the histone/DNA ratio or a conformational change in the nucleosomes associated with genes undergoing active transcription (Weintraub and Groudine, 1976). The chromatin structure of specific genes has also been shown to be conformationally altered only in tissues where they are

expressed. Examples include the β-globin gene in chick embryo red blood cell nuclei and the ovalbumin gene in chick oviduct nuclei (Garel and Axel, 1976). Also, several investigators have proposed that nucleosomes might be "phased" on the chromosome so as to render particular areas of the DNA accessible, or inaccessible, to transcription factors (Gottschling and Cech, 1984; Linxweiler and Horz, 1985). Thus, at this juncture, it became more realistic to assume that the chromatin structure of active genes in comparison to silent loci was a more open and dynamic conformation, yet not necessarily devoid of histones as had been postulated.

In Vitro Transcription

During the early 1970s, several investigators actively pursued the activity (or activities) responsible for the synthesis of the various eukaryotic mRNAs. Almost simultaneously several laboratories were able to isolate multiple RNA polymerase activities on DEAE-Sephadex columns (Chambon, 1975; Roeder, 1976). Each peak of activity exhibited a different susceptibility to the inhibitor amanitin (Kedinger, 1970). There were differences in the results they obtained as evidenced by the diverse number of variant RNA polymerase activities that were originally identified (Roeder, 1976). As the purity of the RNA polymerase activity increased it became more obvious that there were three distinct RNA polymerase activities present in eukaryotic cells (Roeder, 1976). It was very difficult for early investigators to make progress toward understanding the relationship between the various eukaryotic RNA polymerases and their respective function in the

expression of genes, because adequate templates for transcription *in vitro* were not available. The predominant templates used were either homopolymers, bacteriophage DNA, or fractions of genomic DNA enriched in either ribosomal or satellite DNA (Chambon, 1975). These proved unsatisfactory, and the results were often confusing. Several lines of evidence suggested that ancillary factors were necessary in order for RNA polymerase, in particular RNA polymerase II, to exhibit template specific transcription (Chambon, 1975). The application of restriction enzymes to the manipulation of DNA led to the cloning of specific genes that were then suitable as templates for *in vitro* transcription systems (Nathans and Smith, 1975).

The biological implications of the viral model systems that had been studied *in vivo*, and the new DNA cloning technology, prompted several investigators to develop cell free transcription systems. It was obvious that it would be advantageous to work with an *in vitro* system to dissect the various components of the eukaryotic transcriptional apparatus. The first *in vitro* transcription systems were developed for RNA polymerase III, and shortly thereafter, RNA polymerase II. RNA polymerase III is responsible for the synthesis of 5S ribosomal RNA (Ng et al., 1979), tRNAs, and a few viral RNAs including the adenovirus VAI and VAII RNAs (Fowlkes and Shenk, 1980). Cell free transcription of the Xenopus 5S rRNA gene by RNA polymerase III was first demonstrated by Birkenmeier et al. (1978) in nuclear extracts of Xenopus oocytes. At the same time it was shown that cytoplasmic extracts of human KB cells (Wu, 1978; Weil et al., 1979) were able to transcribe selectively cloned 5S rRNA, tRNA, and

adenovirus VA RNA genes. The cytoplasmic extracts were shown to contain a majority of the RNA polymerase III activity (Weil et al., 1979) that had apparently leaked from the nucleus during preparation of the extract. With respect to RNA polymerase II, Manley et al. (1980) prepared a concentrated HeLa cell extract that was able to initiate transcription accurately in vitro at a variety of adenovirus RNA polymerase II transcriptional control regions.

In vitro transcription was and is a powerful technique for the investigation of eukaryotic promoter function. The concomitant development of various molecular techniques for the mutation and reassortment of DNA sequences was fortuitous, and in a relatively short period of time the basic sequence requirements of the RNA polymerase II promoter were delineated (Efstratiadis et al., 1980). Although considerable refinement has occurred in our knowledge of these sequences, the basic elements have not changed. One of the first sequences to be implicated because of similarity to prokaryotic promoter sequences was the "TATAA" box (Goldberg-Hogness). This A-T rich stretch is located -25 to -35 bp upstream of the mRNA start site in RNA polymerase II promoters and is remarkably similar to the Pribnow box (TATAAT) described for the promoters of prokaryotic genes (Pribnow, 1975). The only real difference is the location of the Pribnow box, which is at -10 bp from the start of transcription (Rosenberg and Court, 1979). It should be noted that the comparison of the Pribnow box with the Hogness box has revealed variations in sequence and some difference in function. Principally, the Pribnow box is absolutely required for transcription to occur in prokaryotes; however, as

discussed below, the Hogness box is not as stringently required. The second sequence that has been retained with equally remarkable similarity is the "CAAT" box. The consensus sequence for this element is 5'-GGC$_t$CAATCT-3' (Efstratiadis et al., 1980; Dynan and Tjian, 1985) and is usually located -70 to -80 bp from the mRNA start site.

Although the TATA box and CAAT box have been found in a majority of RNA polymerase II promoters and appear to be the framework around which gene specific variations in regulatory sequences occur, there have been some genes described that have no TATA box (Contreras and Fiers, 1981; Melton et al., 1986; Reynolds et al., 1984). A subset of these genes that have instead a highly G-C rich promoter and in general lack the strict structure created by consensus RNA polymerase II sequences. Examples include enzymes such as mouse dihydrofolate reductase (Farnham and Schimke, 1985), hamster 3-hydroxy 3-methylglutaryl coenzyme A reductase (Reynolds et al., 1984), and human phosphoglycerate kinase (Singer-Sam et al., 1984). These genes are often constitutive and hence have been described as "housekeeping genes." Because the TATAA and CAAT homologies were found in many genes, it was thought that they might function in the regulation of transcription. Early in vitro transcription experiments done by Wasylyk et al. (1980) indicated that the promoter of the conalbumin gene could be deleted to -44 bp from the mRNA start site without any effect on the transcription of the gene. However, when these same investigators introduced even a single base change into the TATAA box, there was a 10 fold decrease in the amount

of transcription. Similar results were obtained with the adenovirus 2 major late control region (Corden et al., 1980; Hu and Manley, 1981; Concino et al., 1984).

In contrast to the *in vitro* results, it was noticed that the TATAA box, in general, was not essential for transcription *in vivo*. Benoist and Chambon (1980) made an SV40 deletion mutant that lacked the TATAA box preceding the early transcription unit. This mutant was capable of synthesizing T antigen and transforming rat cells. Similar results were obtained with the polyoma virus early transcription unit (Bendig et al., 1980). It was also established that the TATAA box preceding the sea urchin H2A transcription unit was not necessary for function *in vivo* (Grosschedl and Birnstiel, 1980a). The deletion mutants that Grosschedl made were assayed by injection into *Xenopus* oocytes. A 54 bp deletion that included the TATAA box lowered the level of transcription 5 fold but did not abolish activity.

If the TATAA box is not absolutely essential *in vivo* for transcription, then what is the function of this highly conserved sequence? The answer came from a series of SV40 early promoter mutants in which the TATAA box was deleted (Gluzman et al., 1980). From this set of mutants it was demonstrated that *in vivo* the initiation of SV40 early transcription occurred downstream of the normal site. Also it was established by Gluzman et al. (1980) that when there were deletions between the start of transcription and the TATAA box the site of initiation remained a constant 25 bp $\pm$ 2 bp downstream. This demonstrated that regardless of the deletion, the mRNA cap site was determined by the position of the TATAA box. Grosschedl and Birstiel

(1980b) found that multiple initiation sites were utilized *in vivo* when the TATAA box was deleted from the sea urchin H2A gene. Since the lack of a TATAA box caused heterogeneity in the start site of transcription for several genes, it is now considered that the TATAA box functions *in vivo* to specify the correct mRNA initiation site.

Early *in vitro* transcription studies did not directly discern whether the CAAT box was necessary for transcription (reviewed in Shenk, 1981). However, more recent and detailed studies have determined that the CAAT box does play a role in transcriptional regulation. Detailed mutagenesis studies by McKnight and Kingsbury (1982); McKnight et al. (1984) and Myers et al. (1986) elegantly demonstrated the need for the CAAT box. Initially the studies of McKnight and Kingsbury (1982), dissected the Herpes Simplex thymidine kinase gene (HSVtk) into discrete areas required for expression: these included the TATAA box and two upstream regions referred to as distal signal I (dsI) and distal signal II (dsII). To pinpoint these small regions accurately they developed a technique called "linker-scanning" mutagenesis which introduces clustered sets of point mutations in a short sequence of DNA. Specifically, these mutations were constructed by ligation of a series of complementary 3′ and 5′ deletions joined via a synthetic linker (BamHI). The mutants that McKnight and Kingsbury created spanned the proximal 120 bp 5′ to the mRNA start site and thus they were able to assign a boundary to all the sequences required for HSV tk gene expression after microinjection into Xenopus oocytes. In subsequent studies dsI and dsII of the HSV tk gene have been shown to interact

specifically with a cellular protein (Jones et al., 1985). This protein, Spl, was initially purified by Dynan and Tjian (1983a) from HeLa cells because of its affinity for the SV40 early promoter--later identified as the G-C rich sequences of the 21 bp repeats. Once the sequence of the binding site (GGGCGG) on SV40 was confirmed by various in vitro methods (e.g., DNaseI footprinting), the purified protein was tested for binding on a variety of other genes that contain a G-C rich sequence(s), including the mouse Dihydrofolate reductase gene (Dynan et al., 1986) and more recently the rat insulin-like growth factor gene by Evans et al. (1988). Both of these genes contain several Spl binding sites, identified in vitro by DNase I footprinting, and the sites in the rat insulin-like growth factor gene are of varying affinity depending on the sequence.

Subsequent to the purification of Spl several groups reported the identification a cellular protein that interacts with the CAAT box sequence and has been referred to as either CAAT box transcription factor (CTF) by Jones et al. (1985) or CAAT box binding protein (CBP) by Graves et al. (1986). Jones et al. (1985) demonstrated an interaction in dsII of the HSV tk promoter between Spl and CTF, thus indicating that distinct transcription factors may interact to regulate expression. The identification of CTF prompted the search for other putative transcription factors, and although the evidence is somewhat preliminary, there appear to be at least 3-4 different CAAT box binding activities depending on the source of the material used to purify the activity and the criteria used for analysis (Dorn et al., 1987). CBP and CTF differ from each other in their heat stability

(McKnight and Tjian, 1986). A CAAT box binding factor isolated from HeLa cells in our laboratory (van Wijnen et al., 1988), termed HiNF-B, is yet another addition to this growing family of proteins, with properties that distinguish it from previous isolated CAAT box-binding factors.

The most sophisticated study to date on the subject of transcriptional regulatory sequences was done recently by Myers et al. (1986). These investigators developed a quick method for the introduction of single point mutations in a small region of DNA. They mutated nearly every base from -1 to -101 bp of the mouse β-globin promoter. With a battery of over 100 clones, each with a single base change in the promoter, they were able to assay the expression of the mutant constructs in vivo in a short term transient assay. Therefore, they could assign functional limits to consensus regulatory sequences and discover any minor, or as yet unnoticed, contributing nucleotides. In addition, transversions and transitions were measured to assess any effects on expression. They demonstrated a requirement for the TATAA box (-25 bp) and the CAAT box (-75 bp) as well as an upstream sequence characteristic of the β-globin genes, CACCC (-96 bp), in β-globin transcription. Significantly, an "up" promoter mutation was discovered when the two bases, GG, immediately 5′ to the CCAAT box were changed to AA. The result of this mutation was a 3-4 fold increase in the level of message. The implications of this result are that a CAAT box transcription factor is able to bind more tightly or more specifically and therefore perform its function more efficiently. With the number of CAAT box binding factors that are being found in various systems, it

is also possible that the "up" mutation results in the binding of an alternative, as yet unidentified, protein that carries out the same function, just more efficiently.

In addition, there are temporal and tissue-specific sequences that are found in the promoters of some genes and regulate expression at the transcriptional level. Many of these elements fall into a category of modulatory sequences referred to as enhancers, negative elements, and silencers.

Enhancers and Silencers

The promoter of a gene has generally been defined as the minimal sequences necessary for the initiation and maintenance of a basal level of specific transcription. Additional elements that modify the expression of a gene either during development, temporally, in a tissue specific manner, or as a result of an inducer, would seem a necessity if adequate regulation in the eukaryotic cell is to be achieved. In the preceding 5-10 years a number of investigators have provided considerable evidence for the existence of positive regulatory sequences referred to as enhancers (Reviewed in Serfling et al., 1985; Maniatis et al., 1987). The properties of an enhancer are that 1) there is strong activation of the linked gene from the correct initiation site, 2) it exhibits independence of orientation, 3) it is operative at long distances whether 3′ or 5′, and 4) it preferentially stimulates transcription from the closest promoter, if they are tandemly arranged (Serfling et al., 1985). The prototype enhancer elements are the 72 bp repeats of SV40, which have been extensively

characterized (Benoist and Chambon, 1980; Fromm and Berg, 1982; Treisman and Maniatis, 1985). Several experiments in which the SV40 enhancer has been fused to the mouse β-globin promoter have demonstrated the relationships that exist between an enhancer and promoter. Banerji et al. (1981) demonstrated that the SV40 enhancer could promote hundred-fold higher levels of rabbit β-globin transcription whether located 1400 or 3300 base pairs away. Treisman and Maniatis (1985) demonstrated that SV40 enhanced transcription of the mouse β-globin gene depended on the presence of a functional promoter. Point mutations in the upstream promoter elements (UPE) of the β-globin promoter abolished transcription almost totally. In conjunction with these results, Treisman et al. (1985) demonstrated that when the β-globin promoter was deleted, and the SV40 enhancer was moved to a proximal position, transcription returned to a high level. It would then appear that enhancers are like promoters but not vice versa. Bienz and Pelham (1986) demonstrated that the tandem duplication of transcriptional control sequences could result in enhancing ability. They found that the duplication of a heat shock regulatory element (HSE) could function as an enhancer (distance activation) whereas a single HSE was inactive at a distance. So one of the major differences between enhancers and promoters (action at a distance) may be due to the number of "promoter" elements present with some accompanying specific sequences (Maniatis et al., 1987). The importance of the specific sequences should not be down-played, as a consensus core sequence, 5'-GTGGAAAG-3', has been identified in viral and cellular enhancers (Khoury and Gruss, 1983).

Differences may also be the result of the arrangement of transcriptional regulatory sequences. Why do an increased number of regulatory sequences in many cases stimulate transcription so dramatically? It has been proposed that the resulting protein-protein complexes that arise from the juxtaposition of regulatory sequences result in increased transcription. Therefore since most enhancers contain repeated elements it is possible that they function in organization of the transcriptional apparatus. Exceptions to this exist of course; tandem duplication of the CCAAT box does not lead to a DNA fragment with enhancer qualities (Bienz and Pelham, 1986), i.e. no enhancement at a distance. Perhaps this result is also a reflection of the idea that some "transcription" factors bind to the DNA but do not act directly. Instead they function through their association with adjacent proteins (Maniatis et al., 1987). An example is that CTF has been shown to associate closely with Spl protein in the Herpes virus tk gene (Jones et al., 1985). Significantly, it has recently become apparent that the mechanism of transcriptional activation by upstream activation sites (UASs) in yeast is conserved in mammals. Several studies over the last year have demonstrated 1) that activator proteins in yeast are composed of a DNA binding domain in the amino terminus of the protein and a transcriptional activator in the carboxy terminus, and 2) that when the yeast proteins are expressed in mammalian cells (with the appropriate binding site present in the promoter of the target gene) they can activate transcription (Kakidani and Ptashne, 1988; Webster et al., 1988; Hope and Struhl, 1986). Taken together with what is known about transcriptional regulation in higher

eukaryotes, it appears that the separation of the DNA binding domain and the transcriptional activation domain of regulatory proteins may be conserved from yeast to mammals. In addition the mechanism is probably conserved as well.

Several of the more well characterized enhancer sequences are part of a group related by tissue specificity of expression. The Immunoglobulin (Ig) enhancer of the heavy chain locus is located several thousand bps 3′ to the variable region promoter. This enhancer sequence, in its entirety, is only active in cells of the lymphoid lineage (Gillies et al., 1983; and Banerji et al., 1983). As has been found for the SV40 enhancer, the Ig enhancer is composed of several distinct elements that interact with specific proteins in vivo (Church et al., 1985). One of the core elements of the Ig enhancer is the "octamer" sequence, 5′-ATGCAAAT-3′. It is of special interest as it also appears in the promoter of a few cellular genes, including histone H2B (Harvey et al., 1982) and (2′-5′) oligo-A synthetase (Benech et al., 1985). How this element contributes to tissue specificity in one context (Ig enhancer) and not in another (histone H2B) remains to be determined. Recent in vitro binding studies of proteins that interact with the SV40 "octamer" sequence have demonstrated that there are both general and tissue specific factors present that bind this sequence, and this may relate to its role in tissue specific regulation (Rosales et al., 1987). Also, careful mapping of the binding of HeLa and B cell nuclear proteins to the SV40 enhancer has revealed subtle differences in the extent to which various motifs are protected which is indicative of differential protein/DNA interactions (Davidson et al., 1986).

Enhancers should not be mistaken for promoters with additional sequence attached or interspersed. In many cases they exhibit exceptional cell-type and temporal specificity with respect to transcriptional activation. Deletion analysis has indicated that certain core sequences of the IgH enhancer may function in non-lymphoid cells to shut off the enhancer action (Wasylyk and Wasylyk, 1986; Kadesh et al., 1986).

The implication of a negative regulatory mechanism for the control of IgH enhancer action presents a confusing picture of tissue specific and temporal gene regulation. At first it was thought that the absence of necessary factors for enhancer action was the reason for differential activity in various tissues (Maniatis et al., 1987). However, this has been shown to be somewhat incorrect as many of the factors found in B cell extracts are also in other types of cells. So, it is either a case of inaccessibility of the DNA binding sites in nonlymphoid cells, or that there must be an interaction with a B cell specific protein (Maniatis et al., 1987). Recently Sen and Baltimore (1986) discovered a factor present in many cell types, NF-kB, that interacts with the kappa-chain gene enhancer, but only after modification to an active form in B-cells.

Negative regulation of gene expression is an old subject for prokaryotic molecular biologists, but is relatively new to eukaryotic gene regulation. The first description of the SV40 enhancer element caused everyone to search for similar elements in other genes, and the identification of negative regulatory sequences, especially in viral enhancers, has had a similar effect. It is important to understand that

negative regulatory sequences can be divided into two groups, 1) those sequences that shut off activity of another regulatory element (such as an enhancer) and have been found to exist within the confines of the enhancer element, and 2) sequences that act independently of other regulatory elements to control the level of gene expression. This latter type of element is the newest discovered and as such is less well characterized. An interesting distinction can be made in that some negative regulatory elements can act in either orientation and with some distance independence and as such have been called either dehancers or silencers (Baniahmad et al., 1987; Laimins et al., 1986; Remmers et al., 1986).

Negative regulation of viral enhancer elements is best typified by the IgH enhancer in which Wasylyk and Wasylyk (1986) have shown that sequences on either side of the central core sequence down regulate expression in fibroblasts as compared to B-cells. It is obvious that, as mentioned above, there must be a mechanism by which the appropriate genes are expressed at the right times in the right tissues. This may occur through the regulation of many protein factors, but more likely there is one protein that regulates the organization of the other transcriptional factors. It seems apparent that the complexity of the eukaryotic promoter would in many cases permit great specificity of expression but could be a regulatory nightmare for the cell. An exquisite example of coordinate regulation of many genes is found in the Adenovirus system and the Ela protein. Ela, one of the immediate early proteins produced in early infection, coordinates the expression

of several other genes (Yee et al., 1987) and also represses the expression of other elements, such as the SV40 enhancer.

A particularly interesting example of negative regulation, which relates to Ela regulation, has been described for embryonal carcinoma cells (EC). SV40, polyoma virus, or Moloney murine leukemia virus are unable to express their genomes when transfected into undifferentiated EC cells. The induction of differentiation removes the block on the expression of both viral and cellular genes (Gorman et al., 1985). Mutants of polyoma virus were isolated that could replicate in the undifferentiated EC cells, and it was found that the mutations occurred predominantly in the promoter and enhancer regions of the early genes. Alternatively, it has been found that the adenoviruses replicate well in undifferentiated EC cells. In conjunction it was discovered that mutants in the Ela region could grow in undifferentiated, but not differentiated EC cells. Taken together with previous evidence about the function of the Ela protein, it has been suggested that EC cells contain an Ela like protein that negatively regulates gene expression until differentiation is induced (Gorman et al., 1985). Gorman et al. (1985) have demonstrated that when the SV40 early promoter is introduced by infection it is inactive in EC cells, but when introduced by $CaPO_4$ transfection it is expressed in an enhancer-independent fashion. This result strongly suggests that the large number of molecules present in the transiently transfected cell are able to titrate out the negative factor (or factors) and thus allow expression from some of the genomes present. Gorman et al. (1985) have also shown that the negative factors in EC cells have different relative

affinities for the various enhancers, and surprisingly the affinity of the interaction did not necessarily relate to the level of expression.

A number of cellular genes have been shown to contain negative regulatory elements although their specific mode of action has not been characterized. These genes include mouse β-interferon (Goodbourn et al., 1986), mouse c-myc (Remmers et al., 1986), rat insulin 1 gene (Laimins et al., 1986), chicken lysozyme (Baniahmad et al., 1987), mouse p53 tumor antigen (Bienz-Tadmoor et al., 1985), chicken ovalbumin (Gaub et al., 1987), and rat α-fetoprotein (Muglia and Rothman-Denes, 1986). This list includes genes in which the negative element is situated within an enhancer (mouse β-interferon) and those in which it is interspersed between other promoter elements (chicken lysozyme and rat α-fetoprotein). The most well characterized of these are the chicken lysozyme and mouse β-interferon genes in which the sequences responsible for the negative effect have been identified (Goodbourn et al., 1986, Baniahmad et al., 1987). The chicken lysozyme gene is particularly of interest because it contains several possible negative regulatory sequences located at -0.25, -1.0 and -2.4 kb from the start of transcription and they are well separated from the enhancer element identified 7 kb upstream (Theisen et al., 1986). Additionally, it is interesting that both the chicken lysozyme and the rat insulin 1 gene negative regulatory elements are contained within repetitive elements. The chicken lysozyme element is found within the CR1 repeat, which is a middle repetitive sequence and has limited homology to the mammalian Alu-type sequences. Additionally, the CR1 repeats near the chicken ovalbumin gene are found in areas where there

is a change in the DNaseI sensitivity when the ovalbumin gene is induced, perhaps indicative of a protein/DNA interaction (Stumph et al., 1984). The rat insulin 1 element is a member of the family of long interspersed rat repetitive sequences (LINES) that are present in about 50,000 copies per cell (Laimins et al., 1986). The fact that some of the negative regulatory elements identified so far are associated with middle repetitive sequences has attracted attention. Some investigators have proposed that the function of this arrangement may be to coordinate transcriptional domains. The isolation of a domain by blocking it off with repetitive elements would be consistent with the structure of eukaryotic chromatin as we understand it today, and would allow for coordinate control of a gene or set of genes of related function (Laimins et al., 1986). Negative regulatory elements are still awaiting the identification of factors that interact with them and characterization of the protein/DNA and protein/protein interactions that result in the negative regulation of transcription.

Histone Genes

Histone proteins have been known for a considerable time and their composition has been the subject of much investigation (reviewed in Isenberg, 1979). Little was known however about the genes encoding these acidic proteins until the late 1960s and the 1970s when many investigators took advantage of the size of the histone messages, and their relative abundance to investigate the regulation of this set of genes. The histone genes have many characteristics that make them an

attractive model system for the investigation of regulation. They are coordinately expressed during S-phase of the cell cycle, and this expression is the result of both transcriptional and posttranscriptional processes. Additionally, their small size and basic structure (no introns, minimal processing) make them an easy system to manipulate and study (Maxson et al., 1983). If we can understand how the highly coupled expression of the histone genes is controlled, perhaps we can then understand how other genes are expressed coordinately and otherwise.

Historical background. One of the initial observations regarding histone proteins was that they are present in a relatively invariant 1:1 molar ratio with DNA in the cell (Prescott, 1966). It was further demonstrated that the amount of histone protein present in a cell doubled during S-phase of the cell cycle (Bloch et al., 1967). Such results suggested a possible coupling between these two metabolic events. Borun et al. (1967) were able to demonstrate that a class of polyribosomes (7-9S) were selectively enriched during S phase of the HeLa cell cycle and that they coded for histone-like polypeptides *in vitro*, thus giving more credence to the relationship that had been demonstrated earlier. Borun et al. also noted several properties of these small mRNAs that have become the foundation of present day theory about histone mRNA regulation: 1) the addition of cytosine arabinoside caused a fourfold increase in the "histone" mRNA destabilization rate as compared to actinomycin D treated cells; 2) the newly synthesized 7-9S RNA, at the G1-S boundary, became associated with polyribosomes thus beginning histone synthesis; and 3) two hours

before the end of DNA synthesis in synchronized HeLa cells 7-9S mRNA transcription ceased and the remaining 7-9S mRNA decayed with approximately a one hour half life. Borun et al. proposed, somewhat incorrectly, that the control of histone mRNA levels was through transcriptional regulation. The refinement of molecular techniques has allowed later investigators to define the degree to which transcriptional and posttranscriptional mechanisms regulate histone mRNA metabolism. Butler and Mueller (1973) repeated and extended the results of Borun by demonstrating several basic facts. First, cycloheximide was able to stabilize histone mRNA in the presence of hydroxyurea, a potent inhibitor of DNA synthesis. When added to synchronized HeLa Cells, hydroxyurea causes a very rapid destabilization of almost all histone mRNAs (90%) via the complete shutdown of DNA synthesis (Baumbach et al., 1984; Heintz et al., 1983; Sittman et al., 1983). This suggests that a protein(s) is (are) necessary for the destabilization process to occur. The 10% of histone message that remains is insensitive to hydroxyurea and probably represents replication independent histone gene mRNAs (Wells and Kedes, 1985; Wu and Bonner, 1982). Second, transcription is not necessary for the production of this putative destabilization factor as the addition of a transcription inhibitor has no effect on the subsequent destabilization of histone mRNA. Third, Butler and Mueller (1973) demonstrated a transient increase in the pool of free histone proteins for 20 minutes after treatment with hydroxyurea. They suggested in their regulatory model that the free histone proteins might autogenously regulate the translation of their own message and/or the

stability of the remaining message following the cessation of DNA synthesis. Nearly 15 years later, the idea of autogenous regulation has gained popularity, since Ross and coworkers (1986, 1987) have so aptly demonstrated the specific degradation of histone mRNA *in vitro*, and the isolation of a nuclease activity that degrades poly A minus messages from the 3′ end.

The histone enriched environment of the sea urchin genome allowed for their early isolation by equilibrium centrifugation and subsequently the characterization of the coding and spacer region base composition (Birnstiel, 1974). The sea urchin genes have been successfully used as probes for the isolation of histone genes from several species, including vertebrates such as *Xenopus* (Moorman et al., 1980) and mouse (Seiler-Tuyns and Birnstiel, 1981). The higher vertebrate histone genes were then used to expedite the isolation of the human histone genes (Clark et al., 1981; Heintz et al., 1981; Sierra et al., 1982). The replication dependent histone genes, which comprise the majority of expressed histone genes, are characterized by a lack of introns and an extremely well conserved 3′ end sequence that consists of an 15 bp stem and loop structure.

Human histone gene organization. The isolation of the human histone genes, which had previously been so intensively studied, permitted the proposed regulatory hypotheses to be tested. The organizational pattern of the human histone genes was uncovered by restriction enzyme analysis, and Southern blot hybridization (Southern, 1975) of restricted phage clones demonstrated that, unlike the tandem repeats of the lower eukaryotes, the human genes were clustered but had no obvious

organizational pattern (Sierra et al., 1982; Heintz et al., 1981 and Clark et al., 1981). Sierra et al. (1982) were able to isolate lambda Charon 4A phage clones representative of three families or clusters. Unlike the lower eukaryotic organization, none of these clustered groups of human histone genes contained a human H1 gene. By using a chicken H1 specific probe Carozzi et al. (1984) isolated a clone that had all 5 human histone genes including an H1 histone. Recently, several human histone genes have been localized to different chromosomes (Triputti et al., 1986, Green et al., 1986). This suggests that coordinate control of human histone gene expression might not be as easily regulated as in lower eukaryotes.

Another question that had not been addressed up to this time was whether different histone mRNAs were the product of different histone genes. Lichtler et al. (1982) demonstrated convincingly that seven species of human H4 histone mRNA were encoded by at least 3 separate genes, thereby establishing that the human histone genes are a repetitive family of genes, but not redundant. Lichtler et al. (1982) also strengthened the possibility that different histone genes might be subject to diverse regulation since it was obvious that certain H4 mRNAs were present at higher levels than others.

Transcriptional and Posttranscriptional regulation. Our knowledge about these two steps in the regulation of histone mRNA metabolism has been strengthened by the studies of Heintz et al. (1983); Sittman et al. (1983) and Plumb et al. (1983a,b). Plumb et al. (1983b) utilized HeLa cells synchronized by double thymidine block and hybrid selection of pulse labelled histone mRNA. This technique permitted several

species of histone mRNA to be isolated on acrylamide gels. These experiments demonstrated that the histone genes are transcribed in the early part of S-phase, approximately 2-3 hours post release from double thymidine block. The increase in the histone mRNA transcription was 3-5 fold during this period. Baumbach et al. (1987) demonstrated a similar increase in the level of histone gene transcription at the beginning of S-phase with nuclear run-on analysis. However, one of the anomalies of histone gene expression is that if one follows the total increase in the amount of histone mRNA, the actual elevation is from 10-25 fold (Plumb et al., 1983b; Heintz et al., 1983). The actual differences in histone mRNA levels have varied from one report to another and this is probably the result of the various synchronization and analysis techniques utilized. Conservatively, the level of transcription increases 3 fold during the first 2-4 hours of S phase, and the stability of histone mRNA rises 10-20 times during S-phase. Outside of S phase or after the artificial cessation of DNA synthesis by drug treatment, the half-life of histone mRNA is approximately 10-15 mins. (Sittman et al., 1983; Plumb et al., 1983a).

Nuclease sensitivity and Protein/DNA interaction. Historically, a hallmark of an active gene has been the presence of nuclease hypersensitive sites in the promoter region of the gene. Chrysogelos et al. (1985) and Moreno et al. (1986) have extensively characterized the nuclease sensitivities of the flanking and coding regions of the FO108 human H4 histone gene. Together, their results demonstrate that the 5' region of the FO108 H4 gene is a dynamic area of varying sensitivity to DNase I, micrococcal and S1 nuclease. Since the histone genes are cell

cycle regulated with respect to transcription and total message levels, Chrysogelos et al. (1985) were able to correlate the size of the DNase I hypersensitive site with the stage of the cell cycle. As mentioned earlier, the appearance of a DNaseI hypersensitive site is indicative of protein/DNA interactions in the region. Pauli et al. (1987) utilized the technique of genomic sequencing to visualize the in vivo protein/DNA interactions in the promoter of the FO108 human H4 histone gene. They demonstrated that there are two binding sites in the proximal promoter region which have been designated Site I (-122 bp to -89 bp) and Site II (-64 bp to -23 bp). Site I contains a putative Spl site and a possible CAAT box. Site II contains the GGTCC element (see below) and the TATAA box. The protein/DNA complexes at Site I and Site II are present throughout the cell cycle and presumably these interactions in the promoter region are involved in the basal and increased level of transcription demonstrated at the onset of S-phase. Perhaps the interactions that regulate the level of transcription at the start of S-phase occur through protein/protein interactions since there is no apparent change in the protein/DNA interactions during the cell cycle. In studies done by Heintz and Roeder (1984), it was demonstrated that the pHuH4 histone gene was transcribed in vitro to a greater extent in S-phase extracts than in G-phase extracts. It would be important to know whether there is a new protein that appears at the onset of S-phase that acts either directly to augment transcription by interacting with the DNA or through a protein/protein interaction. Since the identification of protein/DNA interactions in the promoter of the FO108 H4 gene, it has been of great interest to

us to ascertain if there is any functionality in the interaction and this is addressed to some extent in this work.

Other histone genes, from a variety of species, have been characterized with respect to the contribution of 5' flanking sequences in transcriptional regulation. Notably, the human H2B gene has been extensively characterized with in vitro transcription by Sive et al. (1986). They demonstrated that the transcription of the H2B gene is dependent on a number of sequences 5' to the TATA box including the H2B octamer element and CCAAT box. Recently, the emphasis has been placed on identification of the sequences responsible for the periodic increase in histone gene transcription during the cell cycle. Artishevsky et al. (1987) have demonstrated, although not convincingly, that the sequences responsible for the S-phase increase in transcription of a hamster H3 gene are located in the proximal promoter region (-150 bp); however they were not explicitly defined. The authors propose that this region of the hamster H3 gene bears similarity to the sequence, 5'-CCGAAA-3', that has been shown to regulate the cell cycle expression of the HO genes of yeast (Nasmyth, 1985). Taken as a whole, these many results support the idea that the histone genes are controlled at the transcriptional level by promoters that are composed of many elements that interact with different and specific proteins. Though not dealt with here, van Wijnen et al. (1987, 1988) have shown that the promoter region of several cloned human histone genes can interact with nuclear proteins in a specific manner.

Sequence analysis. Only a few histone genes have been sequenced extensively enough to permit a comparative analysis of 5′ flanking sequences. The majority of sequencing information concerning histone genes has revolved around the coding sequences. Comparative analysis of these protein sequences has revealed remarkable homogeneity from species to species, especially with respect to histones H3 and H4 (Wells, 1986). Unfortunately little 5′ flanking sequence for H4 histone genes has been published, and most sequences extend only 80-120 nucleotides upstream (Wells, 1986). A comparison of the FO108A H4 histone gene (Sierra et al., 1983), which my studies have involved, and the human H4 histone gene independently isolated by Heintz et al. (1981), suggests that some of the sequences in the 5′ proximal promoter region are conserved--the TATA and GGTCC boxes. The TATA box is, of course, a canonical RNA polymerase II transcription sequence and the GGTCC box has been associated with many H4 gene promoters from sea urchin to human (Hentschel and Birnstiel, 1981, Wells, 1986). Comparison of the FO108 gene to the mouse H4 gene isolated by Seiler-Tuyns and Birnstiel (1981) reveals extensive similarity between the promoters, especially the TATA box, GGTCC element, and the CAAT sequence that is found as either a single or double copy located just 5′ to the GGTCC element in many H4 histone genes (Wells, 1986). The significance of the H4 "CAAT" sequence is somewhat questionable as it was originally thought to represent a the "CCAAT" box that is associated with many RNA polymerase II promoters. There have been several CCAAT box factors isolated, and all of them require, for good binding, the sequence 5′-CCAAT-3′ (Dorn et al., 1987). The H4

histone gene with which we are working, F0108, does have two CCAAT boxes located several hundred basepairs upstream and the possible functionality of both the proximal CAAT boxes and the distal CCAAT boxes is discussed in the work presented here.

The functionality of these and other sequences in the promoter of histone genes has been one of the focuses of our work. Also, the Heintz and Roeder laboratory have investigated the functionality of promoter sequences in the human H4 gene they isolated. In vitro transcription analysis of Bal 31 deletion mutants of the F0108 H4 gene by Sierra et al. (1983) demonstrated, in whole cell extracts, that promoter sequences could be deleted to within 50 bp of the cap site without loss of transcription. These sequences include only the TATA box and GGTCC element, but are apparently sufficient for accurate in vitro transcription to occur. In vitro transcription analysis by Hanly et al. (1985) demonstrated very similar effects. When only the TATA box remained as the sole RNA polymerase II consensus element, transcription was accurate but at a reduced level. Hanly et al. (1985) have suggested that the sequences extending to -110 bp are sufficient for maximal transcription of the human H4 histone gene in vitro.

The analysis of histone gene transcription in vitro has contributed to our understanding of the minimal requirements for 5' sequence; however, it has been demonstrated previously that the requirements for initiation of mRNA synthesis in vitro and in vivo are different in many instances. One might reasonably assume that the chromatin structure of

an integrated gene would affect its regulation and intrinsic accessibility to regulatory proteins. We felt it was necessary to extend these in vitro studies into stable cell lines for the reasons outlined above and discussed in Materials and Methods (Chapter 2). A logical extension of many in vitro studies has been to manipulate the promoter or coding region of a gene in vitro and to replace it in vivo and hopefully measure the affect of the manipulation on expression. Perhaps this has been most successfully accomplished in yeast, where the reintroduction of the manipulated gene can be done with precision into the exact locus from which it came originally (Szostak et al., 1983). This is a goal shared by many molecular biologists as it would be a more accurate way to assess structure/function relationships.

Histone genes have been transiently expressed in a number of different cell types (Kroeger et al., 1987; Capasso and Heintz, 1985; Green et al., 1986; Bendig and Hentschel, 1983; Marashi et al., 1986). The transient assay affords a reasonably quick way to examine the effects of DNA manipulation. The results have suggested that heterologous or homologous systems can be used to express transfected genes. In probably one of the more radical transfection experiments, Bendig and Hentschel (1983) introduced the embryonic histone gene repeat of the sea urchin Psammechinus miliaris transiently into HeLa cells. Correct 5' mRNA start sites were detected for all 5 genes of the cluster, but the termination of transcription was generally aberrant with the exception of the H2B gene. This set of results is suggestive that heterologous systems may share many regulatory components that allow them to transcribe foreign genes correctly, but may have--in

this case 3′ processing--parts of the regulatory machinery that are incompatible. This particular subject is discussed in the work presented here. At the point where our work began, the only stable cell lines created with an integrated human H4 histone gene were by Capasso and Heintz (1985). They utilized one construct, pHuH4, to assess the level of H4 histone gene regulation in mouse Ltk$^-$ cells. *In vivo* S1 nuclease analysis of this single construct permitted them to conclude that mouse cells could accurately transcribe the human H4 gene. Green et al. (1986) demonstrated that the F0108 human H4 histone gene was expressed in mouse C127 lung fibroblasts. In these experiments the F0108 gene was carried episomally on a construct made from the 69% transforming fragment of Bovine papilloma virus.

With this understanding and background we initiated studies with the human H4 histone gene F0108 (Sierra et al., 1982) to ascertain the *in vivo* functionality of sequences in the 5′ promoter region.

CHAPTER 2

MATERIALS AND METHODS

Experimental rationale and commentary. Of particular importance, for histone and other eukaryotic genes, is the identification of regulatory sequences and molecules that mediate transcriptional control. Several laboratories, including our own, have conducted *in vitro* and *in vivo* experiments to assess the functionality of the histone gene coding region and flanking sequences in the regulation of expression (van Wijnen et al., 1987; Sierra et al., 1983; Heintz et al., 1983; Pauli et al., 1987; Dailey et al., 1986; Green et al., 1986).

We felt that an *in vivo* approach, via the introduction of modified genes by transfection, had the advantage that the integrated gene was packaged as chromatin and presumably transcription factors, such as RNA polymerase II, CTF, and Spl were present in proper and localized concentrations due to the structural integrity of the nucleus. Therefore the results would be a more accurate reflection of the actual *in vivo* situation. The results were still cautiously interpreted in the context of the experimental parameters present, such as copy number. Some of our experiments have been done in a transient assay system and the expression of the human H4 gene under these conditions was somewhat different than when stably integrated. Presumably there were

differences in chromatin structure and factor to DNA ratios and this may have been reflected in the results. Previous work has demonstrated that the human H4 histone gene, with which we have worked, has a defined chromatin structure that includes an extensive DNaseI hypersensitive site, and that this site fluctuates in size during the cell cycle, which may be the result of the interaction of transcriptional control factors (Chrysogelos et al., 1985).

An in vivo experiment with a transfected gene requires an assay and experimental approach that will allow for the detection of the introduced gene. Several options were available for us to pursue. The most commonly used have been 1) the promoter of a gene was linked to a reporter gene such as chloramphenicol-acetyl-transferase (CAT) (Gorman et al., 1982), or 2) the whole gene, coding and flanking regions, was introduced into a heterologous environment (e.g. a human gene into a mouse cell) (Capasso and Heintz, 1985, Marashi et al., 1986). Several groups, including our own, have utilized such heterologous systems because they allow for the easy detection, by S1 nuclease analysis, of the mRNA of interest with little or no background. We decided that it would be better to leave the H4 promoter attached to the H4 gene and express these constructs in mouse cells.

The histone constructs we cotransfected with the pSV2neo plasmid were expressed and detectable with S1 nuclease analysis in mouse cells. We realized that the histone promoter deletion constructs could be compared to one another and the differences in the steady state level of histone mRNA from one construct to another were a direct reflection of transcription. We concluded this because the coding region of all

the constructs had remained intact. Messenger RNA turnover was presumably the same for each construct and any differences in the steady-state level of histone mRNA were therefore a result of transcription.

We included a mouse H4 control in each of our S1 nuclease assays to permit the quantitation of the total amount of mRNA and particularly the amount of histone mRNA. In retrospect, this has helped us to understand more about the interaction of transcription factors with the H4 histone genes and in some cases has been an adequate internal control. Because of the competition phenomenon we uncovered (described in Chapter 4) the mouse H4 became a less than perfect internal control. Originally we tried to incorporate the mouse 18S ribosomal RNA gene into our S1 nuclease assay but were unable to find adequate hybridization conditions for both histone and ribosomal probes. Ideally another mouse histone gene in conjunction with the mouse H4 should have been used.

Materials and general laboratory procedures. All chemicals were of the highest quality available. Phenol was redistilled and stored frozen with the addition of 0.1 % (w/v) 8-hydroxyquinoline at -20°C. The frozen phenol was equilibrated first with 100 mM Tris-HCl (pH 8.0) and subsequently with 10 mM Tris-HCl and 1 mM EDTA (pH 8.0) until the pH was between 6.0 and 7.0. Phenol/Chloroform extraction refers to the addition of one volume of equilibrated phenol and one volume of Chloroform/isoamyl alcohol (24:1) to a solution, mixing, and separation of the phases by a brief centrifugation step. Next, at least one volume of chloroform/isoamyl alcohol is added and the above

centrifugation step repeated. Hereafter precipitation refers to the addition of 2-3 volumes of 95% ethanol, 1/10th volume 3M Sodium Acetate (pH 5.0), to a solution of DNA or RNA. This was subsequently placed at -20 or -70°C for a sufficient time to allow precipitation of the nucleic acids. Radioactively labelled nucleotides, [γ-^{32}P]ATP (~ 600 Ci/mmol) and [α-^{32}P]dCTP (~ 3000 Ci/mmol), were purchased from Amersham and ICN. X-ray film, Cronex and XAR-5, were obtained from Dupont and Eastman Kodak respectively. For all experiments that involved RNA the solutions were pretreated with 0.01% diethylpyrocarbonate (DEPC) and glassware was treated with 0.1% DEPC. After a 30 min. treatment the solutions and glassware were autoclaved for thirty minutes to remove any traces of DEPC.

Plasmid growth and preparation. L-broth (Maniatis et al., 1982) was prepared by mixing 10g/l Bacto tryptone (Difco), 5 g/l yeast extract (Difco), 5 g/l NaCl,and 2 ml/l 1M NaOH in 1 L of ddH_2O (double distilled water). The medium was then autoclaved for 30 min. in order to sterilize it. Ten milliliter starter cultures of bacteria were prepared in sterile conical tubes and grown overnight at 37°C. These were supplemented with sterile 20% glucose (100 μl), 1M $MgSO_4$ (10 μl), and 50 μg/ml ampicillin (Sigma). Small inocula were removed from glycerol stocks or colonies were picked from plates and placed in the starter culture overnight. Large scale (500 ml) preparations were then completed with 5 ml 20% glucose, 0.5 ml $MgSO_4$ and 50 μg/ml ampicillin. Cultures were grown at 37°C until they reached an optical density (595nm) of 0.4 to 0.5. At this point 4.25 ml of 20 mg/ml chloramphenicol were added and the cultures were allowed to grow for an

additional 16-18 hrs. If the bacteria contained a pUC plasmid or derivative, the amplification step was omitted. The cells were harvested and the plasmid DNA was prepared essentially as described by Maniatis et al. (1982). The pellet was resuspended in 10 ml of Solution 1 (50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA, and 5 mg/ml lysozyme (Cooper Biomedical)) and incubated at room temperature for 5 min. Next, 20 ml of Solution 2 (0.2 N NaOH, 1% SDS) was added and the cells were placed on ice for 10 min. Fifteen ml of Solution 3 (5M KAc, pH 4.8) was added and incubated on ice for 10 min. The cells were then centrifuged at 10k rpm for 20 min., 4°C. The supernatants from all tubes were pooled and precipitated with 0.6 volume of isopropanol for 15 min. at room temperature. The precipitate was recovered by centrifugation at 10k rpm for 30 min. The pellet was dried and resuspended in 8 ml of 10 mM Tris-HCl pH 8.0, 1 mM EDTA (TE). Eight grams of CsCl and 640 μl of 10 mg/ml ethidium bromide were added and the preparation was centrifuged for 36 hrs at 45k rpm in Beckman heat sealed tubes in a Beckman Ti50 rotor. The DNA band was visualized by ultraviolet illumination and recovered by side puncture with a 20 gauge hypodermic needle. The DNA was then either placed over a small Dowex AG 50W-X8 column or butanol extracted 5X to remove the ethidium bromide. The sample was then dialyzed extensively against TE. The DNA was recovered by ethanol precipitation and subsequent centrifugation. Quantitation of the yield was done spectrophotometrically (Beckman) at 260 nm.

Plasmid preparation with TB. The method is similar to that outlined above for L-Broth except that the TB medium was used. TB was prepared

as described by Tartof and Hobbs (1987). Bacto tryptone (6.65 gr.), 13.3 gr. of yeast extract, and 2.2 ml of glycerol were prepared in 450 ml of ddH_2O. The medium was sterilized in the autoclave for 30 min. To the sterile solution was added 55.5 ml of sterile 0.17M KH_2PO_4, 0.72M K_2HPO_4. This medium was inoculated and bacteria were grown as above. Because the medium is very rich, the yields were often large so bacteria that contained pBR322 plasmids were not induced with chloramphenicol. The DNA was prepared by the same method except that the original volume of cells was split into two aliquots at the beginning of the isolation procedure. This was found to be essential and greatly facilitated lysis and subsequent isolation of the plasmid DNA. For comparative purposes, 500 ml of TB can produce 4-5 mg of total plasmid DNA in comparison to 1 mg with L-Broth with amplification.

Production of unidirectional deletions with Exonuclease III. This method was carried out essentially as described by Stratagene (San Diego, CA) from which the reagents were purchased. The method takes advantage of the fact that Exonuclease III cannot digest 3′ single strand overhangs. For our purposes the pFO005 insert was cloned into the PstI/HindIII sites of Bluescript M13+. The HindIII site is adjacent to an ApaI site in the vector. To produce the deletions in which we were interested, the pFO005 Bluescript clone was digested with HindIII (5′ overhang) and ApaI (3′ overhang) to completion. We then mixed three μg of digested DNA, 25 μl of 2X Exonuclease III buffer (100 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 20 μg/ml tRNA), 5 μl of freshly prepared 200 mM 2-mercaptoethanol, 30 units of Exonuclease III, and enough ddH_2O to make

the final volume 50 μl. The reaction conditions were established through a series of titration experiments to determine the extent of deletion with time. After the addition of the enzyme (added last) 10 μl aliquots were removed every minute for 5 min., diluted with 80 μl 1X Mung Bean nuclease buffer (5X = 150 mM NaOAc, pH 5.0, 250 mM NaCl, 5 mM $ZnCl_2$, 25% glycerol) and heated to 68°C for 15 min. Once the deletion reactions had been stopped 9 units of Mung Bean nuclease in dilution buffer (1X = 10 mM NaOAc, pH 5.0, 0.1 mM ZnOAc, 1 mM cysteine, 0.001% Triton X-100, 50% glycerol) were added and the reaction allowed to proceed at 30°C for 30 min. The reaction was stopped by the addition of 100 μl of phenol/chloroform and extracted. The aqueous layer was removed and precipitated with 10 μl of 3 M NaOAc pH 7.0 and 2.5 volumes of 95% ethanol. The DNA was recovered by centrifugation, ligated and transfected as described below. This procedure worked very poorly and resulted in very few positive clones. The deletions that were obtained were characterized by run-off transcription from the T3 promoter of each clone. The DNA was digested with NcoI and transcription reactions carried out exactly as described by Stratagene. The transcripts were electrophoresed on a 6% acrylamide, 8.3M urea gel and the extent of deletion determined by comparison to run-off transcription from the parental construct pFO005BS.

DNA Fragment Elution. After restriction enzyme digestion DNA fragments were usually electrophoresed in low percentage agarose gels (0.7 to 1.0%) with 1X TBE (10X = 500 mM Tris-HCl pH 8.3, 500 mM boric acid, 10 mM EDTA) and visualized by long wave ultraviolet illumination of the ethidium bromide stained band (2 μg/ml for 15 min.). The band of

interest was excised from the gel. The Fragment Eluter (IBI) was first run for 30 min. with low salt buffer (20mM Tris-HCl, pH 8.0; 5 mM NaCl; and 0.2 mM EDTA) at 125 volts. The gel fragment was then placed in the well and the V-channel filled with 100 μl of high salt buffer (3M NaOAc, 5% glycerol, 0.01% Bromophenol Blue). It was important that the gel slice remain in the same orientation as it had been run previously to facilitate the removal of the band. The band was electroeluted at 150 V for 15-20 min. after which the high salt buffer was carefully removed in 100 μl aliquots. A total of 4, 100 μl aliquots were removed from each channel. Five micrograms of glycogen (Boehringer-Mannheim) were added and the sample was precipitated with 1 ml of 95% ethanol at -70°C for 30 min. The DNA fragment was then recovered by centrifugation at 10k rpm for 30 min. Fragments isolated in this manner were found to be directly suitable for ligation reactions or probe preparation.

DNA ligation. The ligation of DNA fragments was done with T4 DNA ligase (New England Biolabs) and essentially as described by King and Blakesley (1986). DNA fragments were digested with the appropriate enzymes dictated by the cloning scheme and fragments and vectors were mixed in 10 μl of 1X ligation buffer (5X = 250 mM Tris-HCl pH 7.6, 50 mM $MgCl_2$, 25% (w/v) polyethylene glycol 8000 (Eastman Kodak), 5 mM ATP, 5 mM dithiothreitol). Usually the vector (a pUC plasmid) was treated with phosphatase prior to the reaction and therefore the vector to insert ratio was ~ 3:1. Blunt end ligations were carried out with less than 20 μg/ml of total DNA. Sticky ligations were done at 20-40 μg/ml and diluted after 4 hrs at room temperature. Generally 10-20 units of ligase were added for sticky end ligations and 200-400 units for blunt

end ligations. After 4 hours the reactions were diluted 1:2 with 1X ligase buffer and an additional aliquot of ligase added to the reaction. The reactions were then incubated overnight at 14°C (sticky end) and 4°C (blunt end). The reactions were diluted 1:2 with TE and transfected into DH5 bacteria as described by the methods of Bethesda Research Laboratories, and Hanahan (1983).

Preparation of competent bacterial cells for transformation. Bacteria, either DH5 or HB101, were grown in 100 ml of Luria broth to an OD_{590} = 0.375. The cells were divided between two sterile 50 ml conical tubes and placed on ice for 10 min. All subsequent procedures were carried out at 4°C. The cells were then harvested by centrifugation for 5 min. at 5k rpm. The supernatant was removed and the cells gently resuspended in 10 ml of $CaCl_2$ buffer (60 mM $CaCl_2$, 10 mM PIPES pH 7.0, 15% glycerol). The cells were then centrifuged for 5 min. at 5k rpm and gently resuspended again in $CaCl_2$ buffer. They were then placed on ice for 30 min. and centrifuged at 2.5k rpm for 5 min. The cells were resuspended in 2 ml each of $CaCl_2$ buffer and dispensed into 200 μl aliquots and frozen at -70°C until needed.

Transformation of bacteria with plasmid DNA. Competent bacterial cells, either DH5 or HB101, were thawed on ice and 5-10 μl of the ligation were added and incubated with the cells for 30 min. on ice. The DH5 cells were heat shocked at 42°C, and the HB101 cells at 37°C. The cells were briefly placed on ice and then diluted with 900 μl of room temperature S.O.C. (2% Bactotryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$). The cells were incubated at 37°C for 1 hour and then plated on TYN (1% Tryptone, 1% yeast

extract, 0.5% NaCl) medium with ampicillin. If detection of insertion of a DNA fragment was possible (DH5 cells and pUC plasmids) then 30 μl of 2% X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) and 20 μl of 100 mM IPTG (Isopropyl-β-D-thiogalactopyranoside) were included with the bacteria spread on the plate. Resistant colonies grew up overnight and white colonies, indicative of a disrupted lac Z gene, were picked for further analysis.

Rapid plasmid preparation. The method is essentially as described by Ish-Horowicz and Burke (1981) with some modifications. One milliliter of saturated overnight culture, grown in TYN or L-broth, was centrifuged for 20 sec. in an Eppendorf microfuge. The solutions for preparation of DNA were the same as for the large scale preparation described above. The cells were resuspended in 100 μl Solution 1 and incubated for 5 min. at room temperature. Solution 2 (200 μl) was added and incubated on ice for 5 min. Solution 3 (150 μl) was added and incubated on ice for 5 min. The cells were then centrifuged for 5 min. and the supernatant extracted with phenol/chloroform. The supernatant was then precipitated with 2 volumes of 95% ethanol at room temperature. DNA was then suitable for restriction enzyme digestion and agarose gel analysis.

Growth and preparation of cell lines. C127 cells were utilized in all transfections and were grown in 10 cm tissue culture dishes as monolayer cultures. The medium used in all experiments was Dulbecco's modified essential medium (Gibco) supplemented with 5% calf serum (Gibco), 5% horse serum (Gibco), 2 mM L-glutamine, and 100 U/ml penicillin, 100 ug/ml streptomycin. To initiate a cell line (histone

plasmid and pSV2neo) or transient (histone plasmid only) transfection the cells were refed with 10 ml of medium 2-4 hours before application of the DNA precipitate. Stable cell lines were initiated by the cotransfection of the histone plasmid and pSV2neo in a 10:1 ratio. This was done essentially as described by Graham and van der Eb (1973) and Gorman et al. (1982). Plasmid DNA, usually 10 μg/construct, was diluted to 450 μl with 1 mM Tris-HCl pH 7.9, 0.1 mM EDTA. This was then mixed with 50 μl of 2.5 M $CaCl_2$. The DNA solution was then added dropwise to 500 μl of 2X Hepes Buffered Saline (280 mM NaCl, 50 mM HEPES, 1.5 mM Na_2PO_4, pH 7.12 $\pm$ 0.05) in a sterile 15 ml conical tube while the tube was vortexed. The precipitates were allowed to stand for 20 min. and were grey and cloudy in appearance. A poor precipitate was obvious as settling out occurred during the 20 min. incubation. The DNA precipitates were added to the plates dropwise under sterile conditions with gentle swirling. After 4 hours the medium was removed and the cells were shocked for 1-2 min with 15% glycerol in medium. This was removed, the cells washed with 10 ml of incomplete medium and refed with 20 ml of complete medium. For transient transfections the cells were incubated for 24-48 hours and then harvested (80-90% confluency) as described below.

Cell lines were initiated by growing the cells to confluency, approximately 2-3 days. At this point the cells were split 1:5 into five plates and the medium was supplemented with 500 μg/ml of Geneticin (G418, Gibco). The aminoglycoside phosphotransferase 3'(II) gene carried on the pSV2neo plasmid confers resistance to this antibiotic and therefore permits cell growth if present. Cells were refed with

medium + G418 every 3-4 days until resistant colonies were apparent and most of the other cells had died. This usually took approximately 2-3 weeks. All the colonies on an individual plate were pooled and subsequently passaged in drug-free medium--these were referred to as polyclonal cell lines. The clone name for a cell line contains several designations. For example: pFO003p1, the pFO designates this construct as originally derived from the λHHG 41 clone isolated by Sierra et al. (1982), 003 describes the deletion construct, and p1 refers to polyclone number 1. When an "m" is used instead of a "p" this indicates a monoclonal cell line. To produce monoclonal cell lines, 12 individual colonies, 2-3 from each plate, were picked with a cotton plugged sterile pasteur pipette and grown in 24 well cell plates (Corning). After these cells had expanded they were grown in 6 and 10 cm dishes as described above.

Cell lines and C127 cells were frozen down periodically in medium supplemented with 20% foetal calf serum (Gibco) and 10% glycerol. Cells were washed off the plate in Puck's Saline + 0.02% EDTA, centrifuged at 1500 rpm for 2 min, resuspended in freezing medium in Nunc Cryotubes, and placed at -70°C.

Southern blot analysis. This method has been used to determine the copy number of the individual monoclonal cell lines and the status of the integrated constructs with respect to flanking sequences and mode of integration. In general, DNAs from individual monoclonal cell lines were digested to completion with restriction enzymes in the buffer recommended by the supplier. The restriction enzyme reactions were stopped by the addition of 1/10 volume of running dye (1X TBE, 50%

glycerol, 0.2% sodium dodecyl sulfate, 0.01% bromophenol blue, and 0.01% xylene cyanol) and heated to 65°C for 15 min. The DNA was then loaded onto 1% agarose gels and run 16-18 hours at 70 V. Gels were stained in ddH_2O with 5 ug/ml ethidium bromide. Next, the gels were soaked in 25 mM HCl for 10 min. to cause strand breaks that permit better transfer and then transferred to Zetabind nylon membranes (AMF-Cuno) as described by Southern (1975) except that the transfer buffer was 0.4 M NaOH (methodology kindly provided by Dr. Harry Ostrer, University of Florida, Department of Pediatric Genetics). Transfer was complete in 20-24 hrs. The filters were gently washed in 2X SSC (20X SSC = 3M NaCl, 0.3M Sodium Citrate, pH 7.0) 3 times for 15 min. each. The filters were briefly air dried and then washed in 0.1X SSC, 0.5% SDS for 1 hr at 65 C. At this point filters were stored at 4°C in plastic Seal-a-meal bags. Blots were prehybridized in 5X SSPE (15X SSPE = 2.69 M NaCl, 150 mM NaH_2PO_4, 15 mM EDTA, pH 7.7), 0.1% SDS, and 1.0% non-fat dry milk (Carnation) at 67-68°C for 4-6 hrs. Hybridizations were performed in the above solution with the addition of either denatured nick-translated or oligolabelled probe. For blots probed with histone H4 sequences 1-2 x 10^6 cpm/ml of probe were used in the hybridization. For mouse 18S ribosomal RNA hybridizations, 1-2 x 10^5 cpm/ml of the pUC974 insert probe were utilized. The specific activity of all probes was at least 1 x 10^8 cpm/ug. The length of hybridization was from 18 - 20 hrs at 67-68°C. Filters were washed 3 times at room temperature with agitation in 5 mM $NaPO_4$ pH 7.0, 2 mM EDTA, and 0.2 % SDS. Each wash was 30 min in length. After a brief drying period the filters were sealed in plastic bags (to prevent dehydration and

facilitate the subsequent removal of probe fragments) and exposed to preflashed XAR-5 film (Kodak) at -70°C.

Preparation of DNA from monoclonal and polyclonal cell lines. The medium from each plate was removed and 2 ml of Puck's saline (Gibco) with 0.02% EDTA were added. The cells were physically removed from the plate by scraping with a rubber spatula and placed in a sterile 15 ml Corex tube. The cells were pelleted by centrifugation at 1500 rpm for 2 min. at 4°C in an IEC-International centrifuge. At this point the supernatant was removed and the cells were snap frozen on dry ice. Frozen pellets were quickly resuspended in 1 ml of 0.1X SSC, 1.0% SDS, and 200 μg/ml proteinase K (Sigma Chemical Company) and incubated for 4 hrs to overnight at 37°C. This mixture was then extracted 2 times and precipitated with 2 volumes of 95% ethanol at -20°C overnight. The precipitated nucleic acids were recovered by centrifugation at 10K rpm for 10 min. at 4°C. The pellet was dried briefly and resuspended in 1 ml of TE and RNaseA (Sigma) was added to a final concentration of 50 μg/ml. Digestion proceeded for 1 hr at 37°C and was stopped by the addition of SDS to 0.5% and phenol/chloroform extraction. DNA was then precipitated with 2 volumes of 95% ethanol, centrifuged at 10K for 10 min, and the pellet resuspended in 500 μl of TE and stored at 4°C.

Copy number analysis. Approximately 30 ug of genomic DNA from an individual cell line were diluted to 50 μl with TE. Digestions were carried out in EcoRI buffer (Boehringer-Mannheim) with the following regime: 1 unit/ug of EcoRI and XbaI were added and incubated at 37°C for 4-8 hrs, at which point an additional 1 unit/ug was added and the digestion proceeded overnight (16-18 hrs). The DNA was quantitated by

diluting 5 μl of the digestion into 1 ml of TE and determining the OD_{260}. The completion of digestion was determined by gel electrophoresis of a small aliquot of the digestion on a 1% agarose minigel (Bio-Rad). Ten micrograms of digested DNA were electrophoresed and blotted as above (Southern Blotting). The probes used for the copy number determination were either the EcoRI/XbaI fragment from pFO002 (for the human H4 histone genes) or the BamHI/SalI fragment from p974 (mouse 18S ribosomal gene for quantitation). The probes were labelled by either nick-translation or oligolabelling (see below). The copy number quantitation of the human H4 histone gene was done by densitometric scanning of multiple autoradiograms. The exact amount of DNA in each lane was determined by reprobing the Southern blots with the mouse 18S ribosomal gene. This gene served as an internal control for variations in the actual amount of DNA loaded and any loss during the process. The copy number of the mouse 18S ribosomal gene should be invariant and all densitometric values for the human H4 histone genes were corrected to account for the actual amount of DNA in the lane based on the internal control.

Labelling of DNA fragments using Klenow fragment. This was done as described by Maniatis et al. (1982). Two hundred nanograms of plasmid or λ phage DNA were digested to completion with the restriction enzymes of choice. One to two microcuries of [α-^{32}P]dCTP were added with ~ 0.5 units of the large fragment of E. coli DNA polymerase I (Klenow fragment, BRL). The reaction was incubated for 10 min. at room temperature. Then 2 μl of 0.2M EDTA, 100 μl of 0.3M sodium acetate, and 20 μg of yeast tRNA were added to stop the reaction. The labelled

DNA fragments were recovered by precipitation with 95% ethanol at -70°C. The DNA was recovered by centrifugation and resuspended in 100 µl of TE.

Nick translation and oligolabelling. Both of these methods were utilized for the production of DNA hybridization probes. Nick translation was done as described by Rigby et al. (1977). For the copy number experiments the EcoRI/XbaI fragment of pFO002 was isolated with the IBI fragment eluter and 250 ng were used in the reaction. A 25 µl reaction was composed of 2.5 µl of 10X buffer (500 mM Tris-HCl pH 7.5, 50 mM $MgCl_2$, 1 mg/ml bovine serum albumin (BSA, Sigma Fraction V)), 2.5 µl of 10X nucleotides (330 µM each of dATP, dGTP, dTTP), 40-80 µCi of α-^{32}P-dCTP, 2.5 units of E. coli DNA polymerase I (BRL), 1 µl of a 1 x 10^{-4} dilution of DNaseI (stored in 10 mM HCl at 1 mg/ml) activated at 1:100 for 1-2 hours on ice in 10 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 1 mg/ml BSA. The reaction was begun with the final addition of the DNaseI and incubated at 14°C for 45 min. The reaction was stopped by dilution with TE and the probe purified over a pipette (10 mm x 100 mm, Fisher) column of Biogel A1.5m in TE. The sample was applied to the column in a 200 µl aliquot and 200 µl fractions were collected. The labelled DNA usually came off in fractions 6-10. These were pooled and quantitated in the scintillation counter. The specific activity of these probes was always greater than 1 x 10^8 cpm/ug. Oligo-labelling was done as described by Feinberg and Vogelstein (1983). The DNA fragment (100 to 200 ng) was added to a 1.5 ml Eppendorf tube and ddH_2O added to make the final volume after addition of the other components either 12.5 µl or 25 µl. This tube was then heated to 95-100°C for two minutes and

placed on ice. To this denatured DNA fragment was added 10 μl of 2X oligolabelling buffer (2X = 500 mM Hepes pH 6.6, 50 μM each of dATP, dGTP, dTTP; 125 mM Tris-HCl pH 8.0, 25 mM 2-mercaptoethanol, 0.55 mg/ml mixed hexanucleotides (Pharmacia)). We added 25-50 μCi of [α-^{32}P]dCTP and 2.5 units of Klenow fragment (BRL). The reaction was allowed to proceed for 2 hours to overnight and purified as described above for the nick translation reaction. Specific activity of these probes usually exceeded 2-4 x 10^8 cpm/μg.

Preparation of total cellular RNA. Because of the sensitivity of histone mRNA to degradation following the cessation of DNA synthesis, it was important that the initial steps of this protocol be carried out as quickly as possible.

The medium from 2-4 plates was removed and 1 ml of cold Puck's saline (Gibco) + 0.02% EDTA was added and the cells were immediately scraped from the dish and transferred to a sterile, DEPC treated, corex tube. The cells were pelleted in the clinical centrifuge at a setting of five for 2 min., the supernatant was removed and the cells were frozen on dry ice and subsequently stored at -20°C for no more than a few days. Degradation can occur quickly and therefore it was necessary to prepare the RNA as soon after harvesting as possible. The cell pellet was resuspended in 1 ml of 2mM Tris HCl pH 7.4, 1 mM EDTA, and 10 μg/ml polyvinylsulfate (PVS, Eastman Kodak). SDS (10%) was added to a final concentration of 1% and proteinase K added to 200 μg/ml. Incubation was at 37°C for 30 min. at which point 5M NaCl was added to a final concentration of 500 mM and the incubation continued for an additional 15 min. The total nucleic acids were extracted with 2

volumes of phenol/chloroform, 2 times, and with 3 volumes of chloroform 1 time. The total nucleic acid was then precipitated by the addition of 60 μl of 3M NaAc and 2.5 vols of 95% ethanol (-20°C overnight). The nucleic acids were recovered by centrifugation at 10K rpm for 15 min. at 4°C. The pellet was resuspended in 500 μl of 10 mM Tris HCl (pH 7.4), 2 mM $CaCl_2$, and 10 mM $MgCl_2$ with the addition of 25 μl of proteinase K treated DNase I (see below for preparation) and digested at 37°C until it was completely suspended (this usually required from 30 min. to 1 hr., intermittent vortexing helped to disrupt the pellet). When the pellet was no longer visible, SDS and NaCl were added to a final concentration of 0.5% and 250 mM, respectively. The solution was extracted 2 times with phenol/chloroform and 1 time with chloroform, and precipitated with 3 vols of 95% ethanol overnight. RNA was either stored in water at -70°C or in ethanol at -20°C. Ethanol suspensions needed to be vigorously mixed to avoid quantitation problems with the RNA aliquots. RNA stored in water was also mixed before removal.

Preparation of RNase free DNaseI. Deoxyribonuclease I (Sigma)(1 mg/ml in 20 mM Tris-HCl pH 7.4, 10 mM $CaCl_2$) was preincubated at 37°C for 20 min. and then further incubated for 2 hrs. at 37°C in the presence of 0.1 volumes of proteinase K (1 mg/ml in 20 mM Tris-HCl pH 7.4, 10 mM $CaCl_2$) to digest any contaminating ribonuclease activity as described by Tullis and Rubin (1980). This preparation was stable on ice for several hours to overnight.

S1 nuclease protection assay. This method is essentially as described by Berk and Sharp (1977) with modifications. In order to

detect the human histone H4 mRNAs 25 μg of total cellular RNA from a C127 cell line containing an integrated human H4 histone gene construct were added to a DEPC treated 1.5 ml Eppendorf tube. Sufficient human and mouse probe, labelled with [γ-^{32}P]ATP, was added to provide an excess (5 to 10 ng) of protected fragment in the reaction. Probe excess was either determined by titration of the probes with a stock C127 or HeLa RNA sample or by addition of twice the amount of probe to some reactions. One twentieth volume of 5M NaCl and 3 volumes of 95% ethanol were added and the solution was placed on dry ice for 15-30 min. The precipitated RNA and probes were recovered by centrifugation at 10K rpm for 15 min. at 4°C. The pellet was briefly dried in a Savant Speed Vac (1-2 min.). Four microliters of 5X hybridization buffer (2M NaCl, 0.2 M Pipes pH 6.4, and 5 mM EDTA) were added followed by 16 μl of recrystallized formamide (Specialty Biochemicals). The buffer was added first to the pellet to facilitate rehydration. The final volume, 20 μl, was vortexed vigorously to resuspend the precipitated RNA and probe. The tubes were placed at 90°C for 10 min. and then transferred immediately to a 55°C water bath and incubated for 12-18 hrs (overnight). Each tube was removed individually from the water bath and the reaction diluted immediately with 8 volumes of ice-cold S1 digestion buffer (280 mM NaCl, 50 mM NaOAc, pH 4.5, and 5 mM $ZnSO_4$) and placed briefly on ice. S1 nuclease (Boehringer-Mannheim) was added to a final concentration of 3 units/μl and digestion was then done at 24-26°C for one hour and at 4°C for 15 min. (the tubes were placed on ice). Ten microliters each of 10% SDS and 5M NH_4OH were added and the reaction was extracted and precipitated with 3 volumes of 95% ethanol.

The length of precipitation was from 3-12 hours at -20°C. (The precipitations should not be done at -70°C as this will cause the formation of formamide crystals). The precipitated probe fragment was recovered by centrifugation at 10K rpm for 30 min. The pellet was briefly dried and resuspended in 2-4 μl of loading buffer (80% formamide, 1X TBE, 0.01% Bromophenol Blue, and 0.01% Xylene Cyanol). Samples were denatured at 100°C for 3 min. and placed immediately on dry ice until loaded. Samples were electrophoresed on a 6% polyacrylamide, 8.3 M urea gel at a 50W constant power for 3-4 hours (the acrylamide to bisacrylamide ratio was 20:1). Gels were dried and exposed to preflashed XAR-5 film (Kodak) at -70°C with Dupont Cronex Lightning Plus Screens.

DNA sequencing. All sequencing reactions were carried out exactly as described by Maxam and Gilbert (1980) and so will not be detailed here. For each fragment that was sequenced the G (Dimethyl Sulfate, (DMS)); G+A (Formic acid); C+T (Hydrazine); C only (Hydrazine in high salt); and A>C (1.2 N NaOH) reactions were done. Single end labelled fragments were prepared as follows: plasmid DNAs were digested with an appropriate restriction endonuclease, treated with phosphatase, and labelled as described below. After the DNA was labelled it was digested with a second restriction enzyme to produce two single end labelled fragments. To purify the fragment of interest for analysis we electrophoresed the DNA on a native 4% acrylamide gel. The location of each labelled DNA band on the gel was determined by exposure to Cronex (Dupont) X-ray film. After alignment of the film and the gel we excised the bands of interest and eluted them in 500 μL of 500 mM ammonium

acetate, 10 mM $MgCl_2$, 0.5% SDS, overnight at 37°C as described by Maxam and Gilbert (1980). The acrylamide gel slice was ground with a siliconized glass rod in a 1.5 ml Eppendorf tube prior to addition of the elution buffer. After the overnight incubation the acrylamide was centrifuged to the bottom of the tube at 10K rpm for 5 min. The supernatant was removed and the pellet resuspended in 200-400 μl of elution buffer, centrifuged, and the supernatant removed. This procedure routinely resulted in recoveries of 80-90% of the labelled DNA fragment. The pooled supernatants were then precipitated twice in succession with 3M Sodium Acetate and 95% ethanol. These fragments were then used in the sequencing reactions noted above. After the reactions were carried out and the DNA was cleaved with piperidine and lyophilized, it was electrophoresed (50W constant power) on a 6% acrylamide, 8.3M urea gel (45 cm x 30cm x 0.5mm). The samples were resuspended in 6 μl of S1 loading buffer and divided into two, 3 μl aliquots. These were boiled for 3 min. and placed on dry ice. To maximize the amount of the sequence we could read, two loadings of the reactions were done. The first 3 μl sample of each reaction was loaded and electrophoresed for 5-6 hours or until the Bromophenol Blue reached the bottom of the gel. The second sample was then loaded and electrophoresed for an additional 5-6 hours. The gel was then dried and exposed to either Cronex or XAR-5 film at room temperature overnight.

S1 nuclease analysis probe preparation. Two probes were routinely used to quantitate the amount of human and mouse histone H4 mRNA present in cell line samples. The human probe was prepared by digestion of 50-100 μg of pFO005 or pFO002 with NcoI. This digestion was then

extracted, precipitated, and the DNA recovered by centrifugation at 10k rpm for 15 min. The pelleted DNA was resuspended in 50 μL of 50 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 1 unit of calf intestinal phosphatase (CIP) was added and the mixture was incubated at 37°C for 30 min. An additional aliquot of enzyme was added and the DNA incubated for 30 min. The reaction was stopped by the addition of EGTA (ethyleneglycol-bis-(β-aminoethyl ether)-N,N,N',N',-tetraacetic acid) to 10 mM and heated to 65°C for 20 min. The DNA was then extracted and precipitated. The DNA was resuspended in 10 μL of γ-^{32}P-ATP (100 μCi) and 1 μL of 10X Kinase buffer (500 mM Tris-HCl pH 7.6, 100 mM $MgCl_2$, 100 mM 2-mercaptoethanol). After resuspension, 15 units of T4 polynucleotide kinase (United States Biochemical Corporation) were added and the reaction incubated at 37°C for 45 min. The reaction was stopped by extraction followed by precipitation. The DNA was recovered, resuspended and digested with HindIII to produce a probe fragment labelled at the NcoI site in the human H4 gene. The reaction was was electrophoresed on a 1.0% agarose gel in 1X TBE and the 695 bp NcoI/HindIII fragment, labelled at the NcoI site purified with the IBI fragment eluter as described by IBI. The mouse H4 probe was produced in a similar manner from the plasmid pBR-mus-hi-1-H4-HinfI (Seiler-Tuyns and Birnstiel, 1981) digested with BstNI. The labelled 1000 bp BstNI fragment was isolated and used as a control in each S1 nuclease protection assay. Although this probe was not single end labelled, we had no ambiguities because of this fact. To make the probe shorter and single end labelled would have possibly obscured the protected fragment of the human H4 gene (280 nt). Both the human and mouse H4 S1 nuclease

probes were quantitated on agarose gels stained with ethidium bromide and exposed to Cronex X-ray film to judge the relative strength of each. Generally a large amount of probe (several micrograms) was prepared simultaneously and S1 nuclease analysis was done on many samples to ensure that the expression was measured with the same strength probe in each case. Variation in the mouse and human probe specific activity did occur; however, the data presented in this work were prepared primarily from a large set of S1 nuclease assays in which many cell lines were assayed side by side with the same mouse and human probe preparation. When additional cell lines were subsequently measured, samples assayed previously were included to ensure that the results could be related to results from previous assays.

Densitometry and data analysis. Densitometry of autoradiograms was done to quantitate the S1 nuclease analysis experiments of H4 gene expression and the copy number of the cell lines. Several films of different length exposure were utilized to determine the intensity of the S1 protected fragment signal. Two densitometers were used, a Zeineh laser densitometer and an LKB-Pharmacia high intensity laser densitometer. Comparison of the capabilities of each densitometer demonstrated that for most films either one was adequate; however for particularly low intensity signals the LKB machine gave more reproducible results. The data collected by both densitometers were computer processed with either the Videophoresis II (Zeineh, Biomed Instruments) or the GelScan XL programs (LKB-Pharmacia). Each program was successfully used to analyze the intensity of radioactive signals for expression and copy number. The areas under the curve for the S1

nuclease analysis (mouse and human) and the copy number blots (H4 and 18S ribosomal) were integrated and expressed as an amount of absorbance units. To calculate the expression of a particular construct, the human expression value was divided by the mouse value and expressed as a ratio. Sample calculations for copy number are presented in Appendix A and for S1 nuclease analysis in Appendix B.

Agarose and acrylamide gel electrophoresis. Agarose (Bio-Rad molecular biology grade) gels were prepared as described by Maniatis et al. (1982). The buffer was 1X TBE and the buffer in the reservoir was also 1X TBE. 20 x 25 cm gels were used for large scale fragment purification and Southern blot analysis of cell line DNAs. Minigels were used for checking the extent of digestion and analysis of rapid and other plasmid preparations. Acrylamide gels were routinely run for S1 nuclease analysis and consisted of 6% acrylamide (20:1 acrylamide to bis acrylamide), 8.3 M urea, and 1X TBE. The gel solution (75 ml) was polymerized with the addition of 750 μl of 10% ammonium persulfate and 20 μl of N,N,N',N',-tetra methylethylenediamine. It was immediately poured, the comb put into place and allowed to harden for 1 hour. Before use the wells were rinsed with buffer and the gel was preelectrophoresed for 30 min. at 50W constant power. The samples were loaded and electrophoresed at 50W constant power.

Genomic sequencing. This technique was done as described by Church and Gilbert (1984). Monoclonal cell lines pFO003ml, 5, and 6 were grown in 15 cm plates (10 per construct). Seven of the 10 were treated with 0.5% DMS in 2-3 mls of medium for 1-2 minutes. Three were left untreated, the DNA purified, and treated with DMS in vitro as a

control. The DMS was removed from the plate and the cells washed twice in phosphate buffered saline (PBS = 150 mM $NaPO_4$, 150 NaCl, pH 7.2, 60 mM Tris-HCl, pH 7.4). The DMS treated cells were scraped from the plate and the DNA purified by incubation with proteinase K as described above and extraction. To purify high molecular weight DNA only, 95% ethanol was slowly added to the tube while swirling the solution with a siliconized glass rod. The DNA was washed off the rod with TE and quantitated spectrophotometrically. The purified DNA (30 μg) was restricted with Hinc II, treated with piperidine and lyophilized as described by the sequencing protocol of Maxam and Gilbert (1980). The samples were then separated in a 6% acrylamide gel, with 8 M urea and electrotransferred to a nylon membrane (Genescreen). The hybridization probe was prepared as described by Pauli et al. (1987) with primer extension of a fragment cloned into M13. In our experiments hybridization was performed with the Hinc II 5′ upper strand probe at 65°C for 16 hrs, followed by eight 5 min. washes at 65°C (1 mM EDTA, 40 mM $NaHPO_4$, pH 7.2, 1% SDS). The membrane was then exposed to preflashed XAR-5 film at -70 C. In these experiments I was responsible for the growth of the cells and Dr. Urs Pauli performed the rest of the experiment, with my constant encouragement, and occasional intervention.

Statistical analysis. The analysis of the S1 nuclease and copy number data that we accumulated was suggested by Dr. Mike Conlon of the University of Florida Biostatistics Unit. After he had examined the data and gained an understanding of the complexities involved, he advised that we employ a ranking test, the Wilcoxon Rank Sum Test. This

test makes the null assumption that two groups of data that are compared came from the same random distribution. The members of each group are assigned a rank (i.e. 1, 2, 3, ...) from highest to lowest in both groups. For example if we had two sets of data, A = 1, 2, 4, 6, and 12 and B = 10, 14, 16, 19, and 25, the members of group A and B would be ranked in order of increasing value. The absolute values of the data are ignored and only the rank is examined.

Group A:(1, 2, 4, 6, 12) is converted to Ranks = 1, 2, 3, 4, 6.

Group B:(10, 14, 16, 16, 25) is converted to Ranks = 5, 7, 8, 9, 10.

We have 5 members in each group with only one point of overlap between the two groups at ranks 5 and 6. The Rank Sum for group A = 17 and for group B = 39. To determine if the difference of the Rank sums is significant, statistical tables of probability for this test were employed. These two groups of data are not significantly different at $p < 0.05$. The reason is the small sample size. With only five members in each group the fact that one of the members of each group falls into the range of the other group precludes any significance. As the groups become larger the overlap allowed for significance becomes greater. I have found with some of my data that larger sample sizes would have been necessary to employ this test in all cases.

CHAPTER 3

HISTONE H4 5′ REGULATORY SEQUENCES

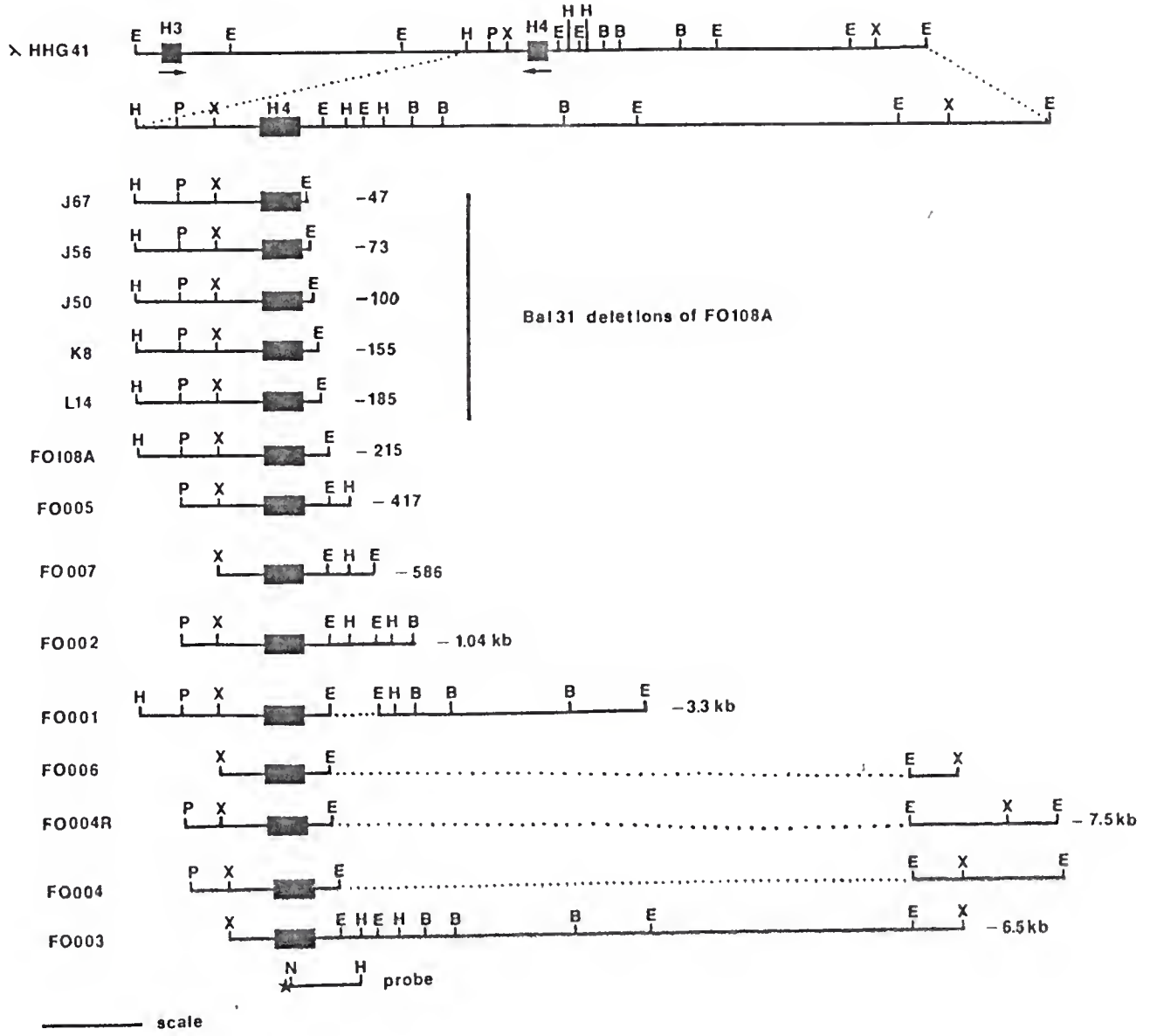
It has been established that the steady state level of histone mRNA during the cell cycle is a function of both transcription and message stability. These two components of histone mRNA metabolism have been studied in a number of different ways. Earlier studies by Plumb et al. (1983a, b) utilized pulsed incorporation of ^{3}H-uridine to determine the contribution of transcription to the increase in histone mRNA levels during the S-phase of the cell cycle. Later, Baumbach et al. (1987) used nuclear run-on transcription to measure transcription of the histone genes directly during the cell cycle. The increase in transcription during early S-phase was determined to be 3-5 fold by both Baumbach et al. (1987) and Plumb et al. (1983b). In the studies of Baumbach et al. (1987), message stability was eliminated as a variable in the experiments, and therefore they were able to determine that histone gene transcription occurred throughout the cell cycle at a basal level. Instead of an "on/off" mechanism for transcriptional control an "enhancement" was apparent during the first 4 hours of S-phase. The 3-5 fold enhancement in the histone gene transcription level has been duplicated in various systems and by different methods during the last 5 years (Sittman et al., 1983; Heintz et al., 1983; Artishevsky et al., 1987).

The implications are that protein/DNA or protein/protein interactions occur that stimulate the increased level of transcription. Evidence for specific protein/DNA interactions has been gathered by Artishevsky et al. (1987). They demonstrated, at the end of G1 and the beginning of S phase, the presence of a factor that interacted with the proximal promoter region of the hamster H3 promoter. The FO108 H4 gene, with which my work has been done, also demonstrates protein/DNA interactions in the proximal promoter region (Pauli et al., 1987, van Wijnen et al., 1987); however, there are no detectable changes in these interactions during the cell cycle. Since it has been demonstrated that transcription of the FO108 H4 histone gene proceeds throughout the cell cycle at a basal level, it was of interest to discover what sequences are necessary for basal and enhanced expression. The promoter of the FO108 H4 histone gene is potentially extensive and so deletions that encompass the entire 6.5 kb of possible promoter sequence were prepared and analyzed. In the proximal region of the promoter we were interested to understand the functionality of elements such as the TATAA box, GGTCC element, Sp1 binding site, and putative CAAT boxes. More distal elements have also been examined and these included a possible enhancer and negative regulatory element located thousands of base pairs upstream.

As mentioned in the introduction, the differences encountered in _in vivo_ and _in vitro_ transcription systems have sometimes been considerable. In order to ascertain the functional _in vivo_ promoter sequences of the FO108 human H4 histone gene, we constructed a series of mouse C127 cell lines each containing a different H4 promoter

Figure 3-1 Schematic diagram of some of the human H4 histone gene deletion constructs.

At the top of the figure is the original λHHG41 phage clone isolated by Sierra et al. (1982). The five Bal 31 deletions of pFO108A are noted (Sierra et al., 1983). The distance from the end of the histone promoter sequence to the cap site is indicated to the right of each construct. FO001, FO006, FO004, and FO004R are fusions of the proximal promoter region and coding sequences to distal fragments and the dotted line indicates the extent of the deletion that occurred between the two fragments. The scale at the bottom is 2 kb on the λHHG41 schematic and 1 kb on all others. The pertinent restriction enzyme sites are denoted EcoRI, E; BamHl, B; HindIII, H; XbaI, X; NcoI, N. The most commonly used S1 nuclease probe is designated at the bottom of the figure labelled at the NcoI site.



Bal31 deletions of FO108A

deletion construct. As described in the prologue to the Materials and Methods section, we decided that this was the best way to proceed. We hoped that stable integration into the chromosome would give the most accurate information about the function of H4 promoter sequences.

Cell line construction

The first step in these experiments was to construct the cell lines. The mouse C127 cell line was chosen because it was a heterologous host and had been previously used to support the stable expression of the FO108 human H4 gene in an episomal form (Green et al., 1986). Many of the histone H4 plasmid DNA constructs were available already (Figure 3-1), although as the work progressed several more were prepared to answer various questions that arose. The constructs are all products of subclones of the original λ human histone gene clone 41 (λHHG41) isolated by Sierra et al. (1982) and this is diagramed at the top of Figure 3-1. The proximal deletion constructs J67, J56, J50, K8, and L14 (Figure 3-1) were all available and had been made by Bal31 deletion of pFO108A (Sierra et al., 1983). The precise determination of each deletion point will be outlined later in the chapter. A subclone of pFO108, pFO108A, prepared by Sierra et al. (1983) deleted some 3′ sequences including an Alu repeat. Plasmid pFO005 was made by A. van Wijnen from a HindIII digestion of pFO002. Plasmid pFO002 was prepared from a BamHl, PstI digest of λHHG41 to obtain a fragment with 1065 bp of 5′ flanking sequence. Plasmid pFO003 was prepared from an XbaI digest of λHHG41 and has 6.5 kb of 5′ flanking sequence. Additional clones will be described as they pertain

to subjects under discussion later in the chapter--positive and negative regulatory elements.

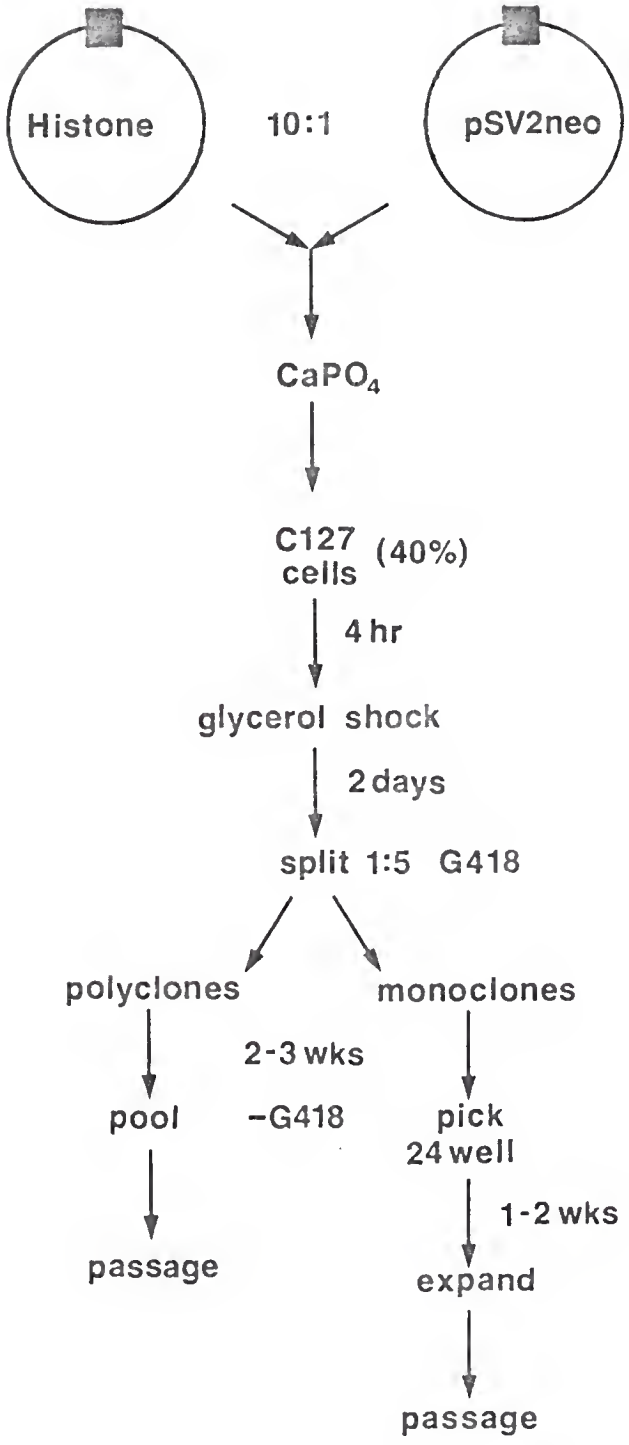
Initiation of Transcription and Basal Regulation

The initiation of transcription by RNA polymerase II and the sequences required for it have been studied in considerable detail in a number of genes, as outlined in the introduction (Reviewed in Shenk, 1981). The importance of the TATA box has been established in vitro and in vivo, and it is thought to be primarily responsible for the specification of the transcription initiation site. We constructed cell lines with several of the short proximal deletion constructs in order to ascertain what sequences in the FO108 H4 histone gene were necessary for the initiation of transcription. The general protocol for DNA transfection and the subsequent selection and expansion process is outlined in Figure 3-2. The constructs were cotransfected into C127 cells with the plasmid pSV2neo. The inclusion of the pSV2neo plasmid permitted selection for expression with the antibiotic Geneticin (G418). Once resistant cells were present as distinct colonies the plates were either pooled and passaged (polyclones) or picked and expanded as monoclonal cell lines. The specific method is described in the Materials and Methods section.

To determine the level of transcription from each of the proximal deletion constructs, we analyzed cell lines early in passage. The results from S1 nuclease analysis of total cellular RNA from polyclonal cell lines 108A, L14, K8, J50, J56, and J67 is presented in Figure 3-3. RNA was prepared from each cell line as described and hybridized to two probes, human and mouse, at 55°C for 8-16 hours as described in

Figure 3-2 Flow diagram for the production of both polyclonal and monoclonal mouse cell lines that contain stable integrated human histone H4 genes.

The method relies on the cotransfection of the histone plasmid with a selectable marker, pSV2neo. This plasmid carries the gene that confers resistance to a derivative of neomycin. The cotransfection procedure permitted the pSV2neo plasmid to be taken up with the histone plasmid into the mouse C127 cells. These stable cell lines were utilized to study human H4 gene expression. The specific protocol is outlined in materials and methods.


Histone
10:1
pSV2neo
CaPO4
C127 cells (40%)
4 hr
glycerol shock
2 days
split 1:5 G418
polyclones
monoclones
2-3 wks
-G418
pool
pick 24 well
1-2 wks
passage
expand
passage

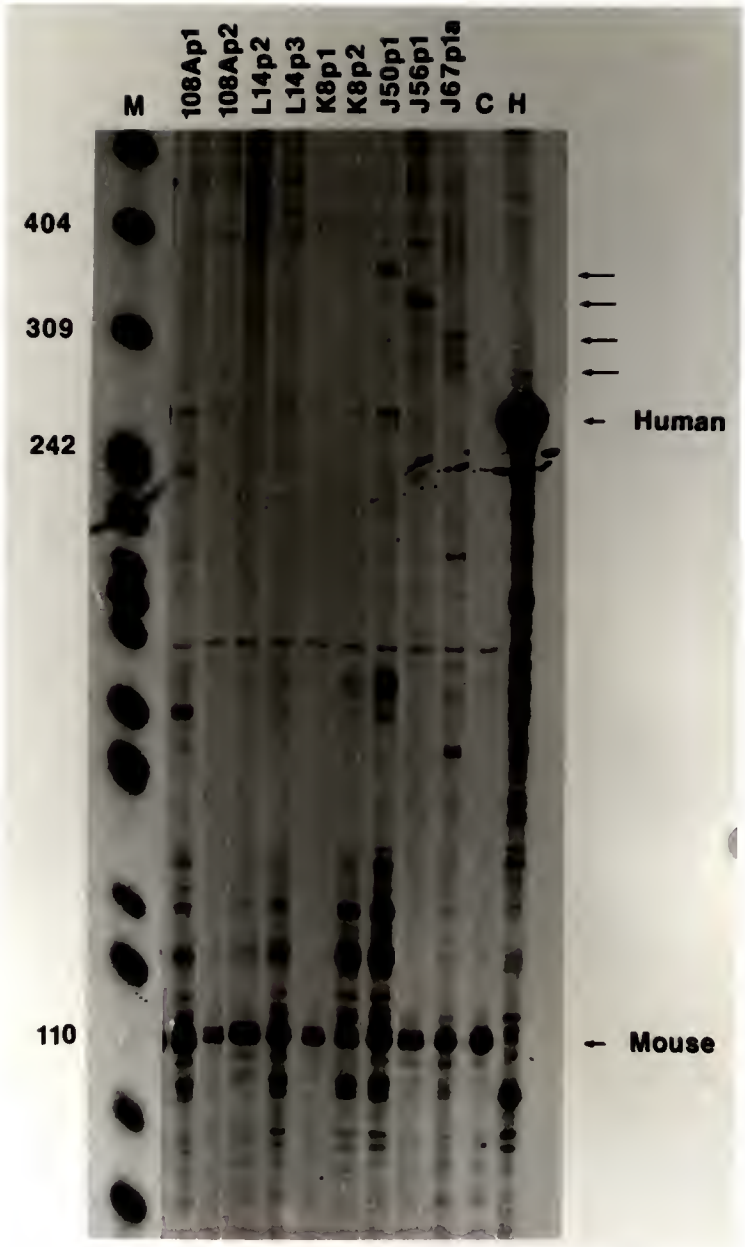


Figure 3-3 S1 nuclease analysis of proximal deletion polyclonal cell lines.

S1 nuclease analysis was done as described in Materials and Methods and quantitated by densitometry. Lanes: the cell line name is denoted above the lane. For example polyclonal cell line pFO108A number 1 is denoted as 108Ap1; C, C127 total cellular RNA and H, HeLa total cellular RNA incubated with both human and mouse S1 probes as a positive control for the size of the mouse and human S1 protected fragments, respectively; M, pBR322 HpaII marker labelled with α-^{32}P-dCTP and Klenow fragment. Both human (280 nt) and mouse (110 nt) protected fragments are noted at the right.

Materials and Methods. The mouse H4 histone probe was included as an internal control in each S1 nuclease assay not only for the intactness of the RNA preparation, but also as an indicator of the amount of histone mRNA present in the sample. The half-life of a histone mRNA after the cessation of DNA synthesis is very short (Plumb et al., 1983a, Sittman et al., 1983), and therefore the growth conditions of the cells and temperature at the time of harvested are critical for the adequate recovery of histone mRNA.

We particularly wanted to determine if there was a minimal amount of promoter that could initiate transcription in vivo and if this was different than that seen in vitro. Previously the shortest Bal31 deletion, J67, had been shown to initiate mRNA synthesis accurately in vitro in a whole cell extract (Sierra et al., 1983). As shown in figure 3-3, the construct J67, which we later learned has only the TATA box and the GGTCC element, produced no correctly initiated transcripts. The only transcription products detectable from the J67 construct were inititated upstream of the normal mRNA start site. These are denoted with arrows in Figure 3-3, and occur in the cell lines with J50, J56, and J67 integrated. The upstream transcription start sites map primarily to the TATA box (-30 bp) and the deletion end points. The "deletion end point transcripts" originate from outside of the histone flanking sequences either in the plasmid or surrounding chromosomal DNA and are detected by virtue of the lack of homology between the probe and the mRNA past the deletion point.

The possibility that J67 was unable to express correctly initiated H4 mRNA was based on a single polyclonal cell line. To assure ourselves that this was not a result of a spurious integration event we

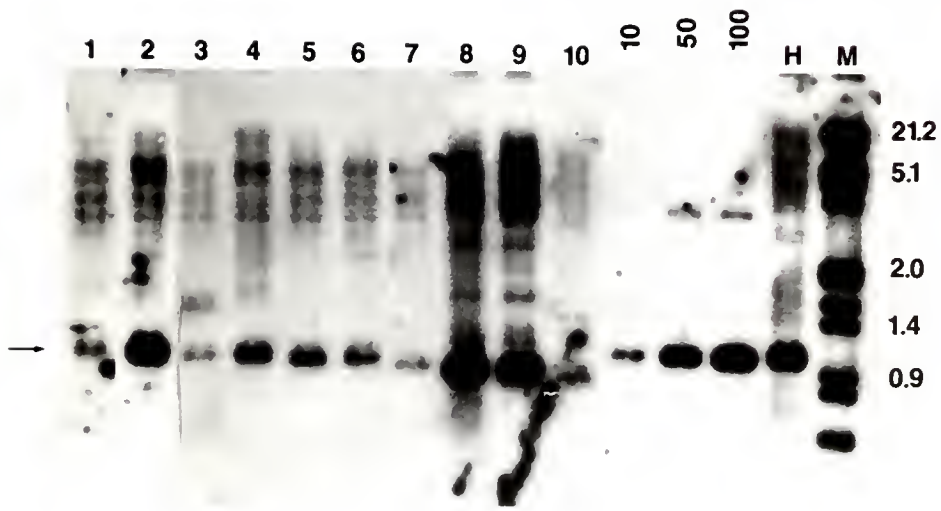


Figure 3-4 Southern blot analysis of polyclonal cell lines: Intactness of 5′ flanking regions and copy number of the constructs in each cell line.

Genomic DNA purified from each cell line was digested with EcoRI and XbaI, electrophoresed, blotted, probed, and quantitated as described in Materials and Methods. Lanes: 1, pFO108Ap1, 2, pFO108Ap2, 3, L14p2, 4, L14p3, 5, K8p1, 6, K8p2, 7, J50p1, 8, J56p1 (passage 4), 9, J56p1 (passage 8), 10, J67p1a. Histone plasmid markers (EcoRI/XbaI digested pFO002) were included on the blot equal to 1.3 (10 pg), 6.5 (50 pg), and 13 (100 pg) gene equivalents per diploid genome in order to quantitate the human histone H4 copy number. H, HeLa DNA digested with EcoRI and XbaI as a positive control for the 1070 bp fragment. M, λ DNA digested with EcoRI and HindIII and Klenow labelled. Pertinent sizes are denoted to the right in kilobases. The probe for this experiment was the EcoRI/XbaI fragment of pFO002 that had been nick-translated as described in Materials and Methods.

determined the intactness of the flanking and coding sequences for each of the constructs J67, J56, J50, K8, L14 and 108A in Figure 3-4. This experiment also permitted us to determine the copy number of each cell line. Ten micrograms of genomic DNA from each cell line was digested to completion with EcoRl and XbaI and electrophoresed on a 1% agarose gel, blotted and probed as described in Materials and Methods. In order to quantitate the copy number of each cell line the gel also contained plasmid DNAs of known amounts digested with both EcoRI and XbaI. Ten, 50 and 100 pg correspond to 1.3, 6.5 and 13 gene equivalents per diploid genome respectively as designated in Figure 3-4. Several exposures of the autoradiogram were scanned with a Zeineh laser densitometer and quantitated in comparison to the controls. Additionally, the Southern blot in Figure 3-4 was quantitated for the actual amount of DNA by densitometrically scanning a photographic negative of the gel prior to transfer, and differences in DNA amounts have been taken into account in the copy number calculation. Later, copy number blots for other constructs were reprobed with a clone of the mouse 18S ribosomal gene kindly provided by the Dr. David Schlessinger (Washington Univ., St Louis) to allow exact determination of the amount of DNA loaded in each lane and subsequently transferred. A sample copy number calculation in which the ribosomal probe was utilized is presented in Appendix A.

The Southern blot analysis demonstrated not only the copy number of each cell line, but permitted us to conclude that the flanking region of most constructs was intact. The mode of integration for the histone plasmids is described further in chapter 4.

Table 3-1 Quantitation of Polyclonal Cell Line Expression.

Cell Line	Human/Mouse Exp	Copy number	Exp/Copy number
108Ap1	0.016	1	0.016
108Ap2	0.040	13	0.003
L14p2	0.018	1	0.018
L14p3	0.017	4	0.004
K8p1	0.028	3	0.009
K8p2	0.029	2	0.014
J50p1	0.034	1	0.034
J56p1	0.019	50	0.0004

A quantitative summary of the expression data from the polyclonal cell lines of the proximal deletion constructs. The human/mouse expression ratio was determined by densitometry of the S1 nuclease protected fragments in Figure 3-3. Copy number for each cell line was determined from the Southern blot in Figure 3-4. Since these data were derived from polyclonal cell lines it is not possible to interpret the results strictly, and we would like to note that copy number in a polyclonal cell line is somewhat ambiguous. Expression is denoted as Exp.

The results of the S1 nuclease analysis and copy number determination are presented in Table 3-1. The S1 nuclease assay was similarly quantitated with the densitometer and the results are expressed as a ratio of the mouse and human signals. The results, although of a few individual cell lines, have been repeated several times. The S1 nuclease analysis results from the proximal deletion polyclones suggested that J67 (-47bp) was unable to correctly initiate histone mRNA transcription. Only when the promoter was extended in J56 (-73 bp) was correct initiation observed (Figure 3-3). It can be seen from the data in Table 3-1 that the expression per copy of the J56 construct (-73 bp) is quite low in vivo (expression/copy = 0.0004), and as noted later this may be somewhat a reflection of the copy number and not the amount of 5' sequence present in the construct. When the flanking sequences are extended to -100 bp in the construct J50 there is an apparent 80 fold increase in the expression/copy ratio (0.034). The expression/copy ratio of the remaining deletion constructs stabilizes at a value of 0.02 to 0.01 with increased length of 5' sequence. This 25-50 fold increase is probably exaggerated because of copy number differences between J56 and the longer constructs. This phenomenon (expression versus copy number) will be discussed later in the chapter. Still it is likely that the difference in the expression/copy ratio is 10 fold. These data are supported by the results of Ken Wright in our laboratory, who has utilized in vitro transcription to define the functionality of proximal promoter elements and demonstrated that in nuclear extracts the transcription of J50

Figure 3-5 Schematic diagram of the proximal human histone H4 Bal31 deletion mutants: Sequence analysis of the deletion points.

Each construct was sequenced according to the protocol of Maxam and Gilbert (1980) and as described in Materials and Methods. The deletion point of each construct is denoted with an asterisk over the last nucleotide included in the sequence of that construct. For reference the ATG codon, TATA box, GGTCC element, CAAT boxes and Sp1 site have been underlined. The two bolded regions of the promoter correspond to Site I and Site II, the DNAseI protected regions of protein/DNA interaction as defined by Pauli et al. (1987).

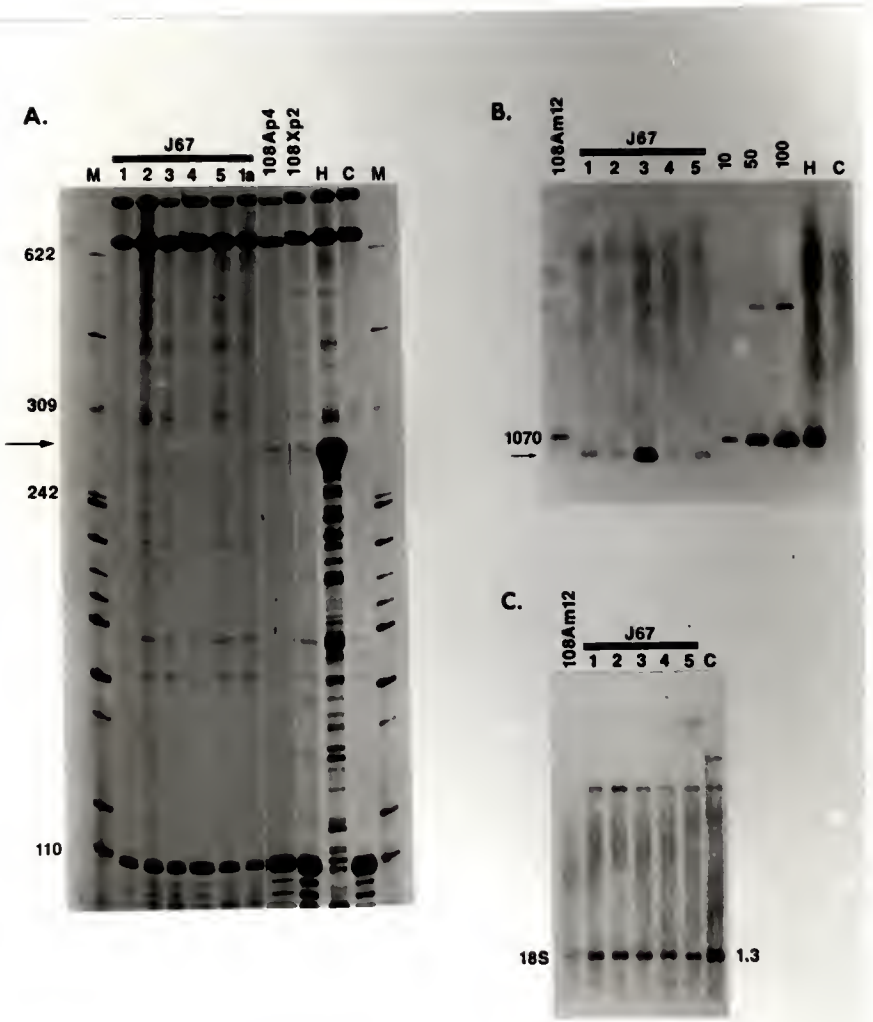
(-100 bp) is several fold higher than J56 (-73 bp) (Ken Wright, personal communication).

Previously, the deletion points of the Bal 31 deletions had been determined by restriction enzyme analysis and electrophoresis on high percentage agarose gels (Sierra et al., 1983). To determine exactly the deletion point, each construct was sequenced by the method of Maxam and Gilbert (1980). Ken Wright and I collaborated in this effort and the approach we undertook is described in Materials and Methods. Importantly, the strategy permitted us to sequence across the deletion point in each construct and to determine the exact end of Bal31 digestion. The deletion points we determined are denoted in Figure 3-5.

When we examined the sequence of the J67 (-47bp) deletion, it was obvious that the GGTCC element and TATA box were still present and the proximal CAAT box (-53 bp) was absent. Our S1 nuclease analysis suggested that this was not sufficient promoter sequence for correct in vivo transcription initiation. To ensure that this was indeed the case, we prepared 5 additional polyclonal cell lines of J67 and demonstrated that they all contained integrated constructs (Figure 3-6b,c); however, none expressed a correctly initiated histone H4 mRNA (Figure 3-6a). The absence of a detectable S1 protected fragment in the J67 polyclonal cell lines was repeated several times. Upstream initiation of transcription was sometimes detectable although this was not consistent. The importance these results became apparent when Drs. Urs Pauli and Susan Chrysogelos of our laboratory demonstrated the binding of proteins to the proximal promoter region of this H4 gene in

Figure 3-6 S1 nuclease and Southern Blot analysis of J67 polyclonal cell lines for correct human H4 expression and copy number.

Additional J67 polyclonal cell lines were made to confirm that this construct was unable to initiate human H4 mRNA transcription correctly. A. S1 nuclease analysis of 25 μg total cellular RNA from 5 new J67 polyclonal lines and the one tested previously, J67pla. Also shown are polyclonal lines 108Ap4 and 108Xp2. H, HeLa total cellular RNA. C, C127 total cellular RNA. M, pBR322 HpaII markers. The human H4 S1 protected fragment (280 nt) is noted with an arrow at the left. There was no detectable human H4 signal in any of the J67 lanes even upon repetition and long exposure. B. Southern blot analysis of J67 polyclonal cell line for copy number determination. J67 polyclones 1-5 and pFO108Aml2 are shown. The position of 1070 bp is noted and the arrow indicates the size of the deletion EcoRI/XbaI fragment from J67. Plasmid DNAs in the amount of 10, 50, and 100 pg were included for copy number quantitation as described in Fig 3-4. H, HeLa cell DNA digested with EcoRI and XbaI; C, C127 cell DNA digested with EcoRI and XbaI. C. The blot in B was reprobed with the 18S mouse ribosomal fragment for quantitation of the amount of DNA in each lane. The size of the 18S band, 1.3 kb, is noted at the right. Quantitation was done as described in Materials and Methods and Appendix A.


A.
J67
M 1 2 3 4 5 1a
108Ap4
108Xp2
H C M
622
309
242
110
B.
108Am12
J67
1 2 3 4 5
10
50
100
H C
1070
C.
108Am12
J67
1 2 3 4 5 C
18S
1.3

vivo (Pauli et al., 1987). The specific areas of protein/DNA interaction as defined by DNase I protection are outlined in Figure 3-5 with the construct deletion end points. Interestingly, the J67 deletion point is located in the middle of Site II and leaves the proximal portion with the GGTCC element and TATA box intact. It would appear that the absence of Site I and the presence of only half of Site II are insufficient for transcription initiation in vivo. However, when all of Site II is present in the of construct J56 a low but detectable level of transcription is present (Figure 3-3 and Table 3-1). The large increase in the expression/copy ratio of the J50 (-100 bp) construct is apparently the result of remarkable similarity to the Spl (Dynan and Tjian, 1983b) binding site as described by Briggs et al. (1985) and Evans et al. (1988). Although we have not proven that the protein/DNA interaction at this site is the result of Spl, it seems a strong possibility that it could be Spl or a similar protein. J50 also includes a putative CAAT box, however the functionality of this sequence is in question because it lacks the necessary homology to the consensus sequence. Additionally, this CAAT box is not entirely included in the protein binding domain of Site I as described by Pauli et al. (1987) and it is therefore unlikely that it functions in the same capacity. It should be mentioned that Spl has been shown to interact with CTF in the HSVtk promoter (Jones et al., 1985), and possible interaction in the histone promoter should not be ruled out immediately, however it is unlikely. The CAAT sequence is well conserved evolutionarily in conjunction with the GGTCC element (Wells,

1986) and our results suggest that the removal of this element in the distal half of Site II prevents correct transcription initiation.

We investigated the whether any diatl promoter elements had an effect on the transcription of the FO108 human H4 histone gene. Polyclonal cell lines were prepared from constructs pFO005 (-417 bp), pFO004 (-6.0 to -7.5 kb), pFO002 (-1065 bp), and pFO003 (-6.5 kb). The results of the S1 nuclease analysis and limited copy number analysis on these cell lines suggested that upstream sequences beyond those already examined might contribute to an increased level of expression (data not shown). Upon reflection, it is likely that in most cases, the increased level of expression we noted was the result of high copy number, and not necessarily because of a strong promoter sequence such as an enhancer. These results, although limited at the time, prompted us to examine in a more rigorous way the distal 5' promoter sequences of the FO108 H4 histone gene for possible regulatory areas that control expression.

Transfection of the constructs pFO005 (-417 bp), pFO002 (-1065 bp), and pFO003 (-6.5kb) into mouse C127 cells was done to assess any distal contributions to the expression level of this H4 gene. As stated previously enhancer and silencer/negative regulatory elements can be located at considerable distances from the promoter of a gene and still accentuate or depress expression of the linked gene (Maniatis et al., 1987, Theisen et al., 1986, Baniahmad et al., 1987). The new cell lines were grown primarily as monoclones, and for continuity with the previous studies, monoclonal cell lines of pFO108A and K8 were also prepared.

I will state now that we have found that there is a competition between the transfected human H4 histone genes and the endogenous mouse H4 gene for regulatory factors and this is discussed later and in chapter 4. The interpretation of expression from each construct is affected by this competition phenomenon, and becomes rather confusing. We bring this up here only to make the reader aware that this situation exists, and the results have been interpreted several ways, sometimes with this taken into account. It has been extremely difficult to understand the relationship that exists between the endogenous mouse H4 genes and the transfected human H4 genes. We have analyzed the expression/copy data carefully to decipher any trends. The results of this analysis are also reviewed in chapter 4. The choice of the mouse H4 as an internal control for the S1 nuclease analysis was both fortunate and detrimental to our interpretation. In short, the entire expression analysis is presented here, but because of the realization later in the course of this work about copy number and competition for transcription factors, only some of the data will be incorporated into the final synopsis.

The monoclonal cell lines were analyzed for the level of expression and copy number present. The S1 nuclease analysis of the pFO003 monoclonal cell lines is presented in Figure 3-7 and was done as described in Materials and Methods. Almost all of the monoclones were positive for expression of the human H4 histone gene with the exception of pFO003m18. We utilized several exposures to determine, densitometrically, the level of expression from each cell line. The expression data are presented as a ratio of the human and mouse

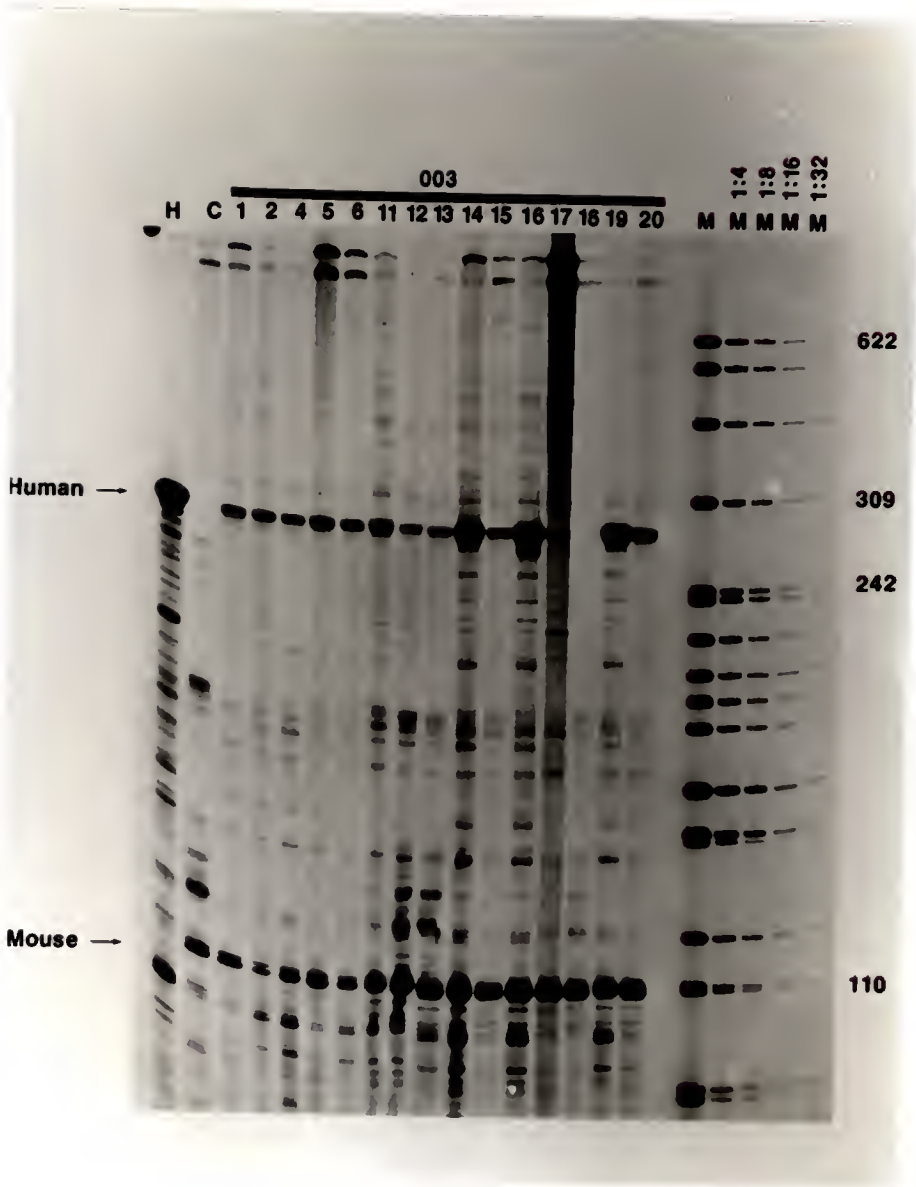
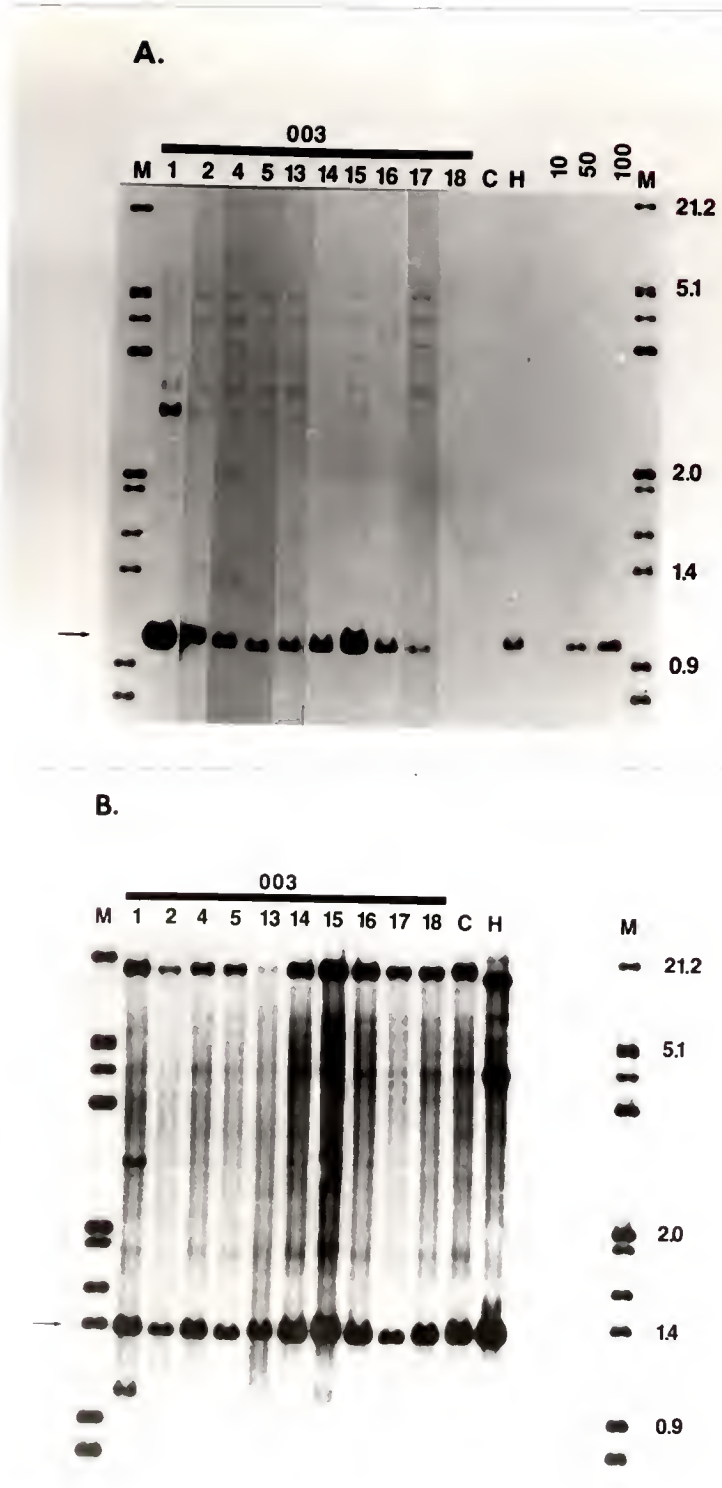


Figure 3-7 S1 nuclease analysis of pF0003 monoclonal cell lines.

S1 nuclease assays were performed as described in Materials and Methods. Almost all 15 clones shown here are positive for expression of the human H4 gene. The exception is pF0003m18. H, HeLa total cellular RNA. C, C127 total cellular RNA. M, pBR322 digested with HpaII and labelled with α-^{32}P-dCTP and Klenow fragment. Dilutions of the marker are noted as 1:4, 1:8, 1:16 and 1:32 for densitometry purposes. The human (280 nt) and mouse (110 nt) protected fragments are denoted with labels and arrows at the left. The clone numbers appear above the individual lanes to which they correspond.

Figure 3-8 Southern blot analysis of pFO003 monoclonal cell lines.

Southern blot analysis was performed as described in Materials and Methods. 10 μg of DNA from each cell line were analyzed with nick translated EcoRI/XbaI fragment from pFO002. A. pFO003 cell line DNA probed with H4 sequences. B. The histone probe was removed and the blot was reprobed with the mouse 18S ribosomal fragment. Densitometry of the 1070 bp band specified by the arrow in A and the 18S ribosomal band in B permitted quantitation of the copy number through normalization to the amount of DNA actually loaded and transferred as described in the Materials and Methods. The figure in A is a composite of several exposures that reflects the actual copy number and accounts for original quantitation errors. The plasmid controls for quantitation are labelled 10, 50 and 100 designating the number of pg loaded. C, C127 cellular DNA. H, HeLa cellular DNA. M, λ DNA digested with EcoRI and Hind III and labelled with α-^{32}P-dCTP and Klenow fragment. The number of each clone is designated above the lane.


A.
003
M 1 2 4 5 13 14 15 16 17 18 C H 10 50 100 M
21.2
5.1
2.0
1.4
0.9
B.
003
M 1 2 4 5 13 14 15 16 17 18 C H
M
21.2
5.1
2.0
1.4
0.9

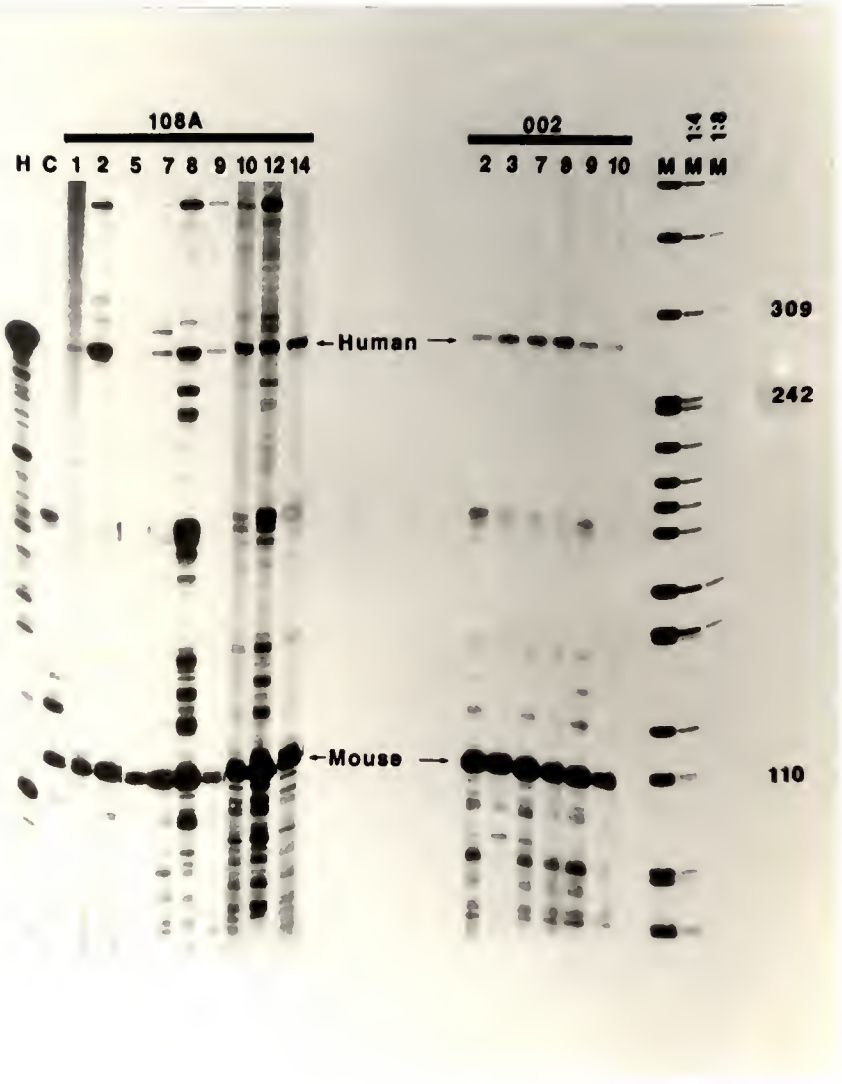
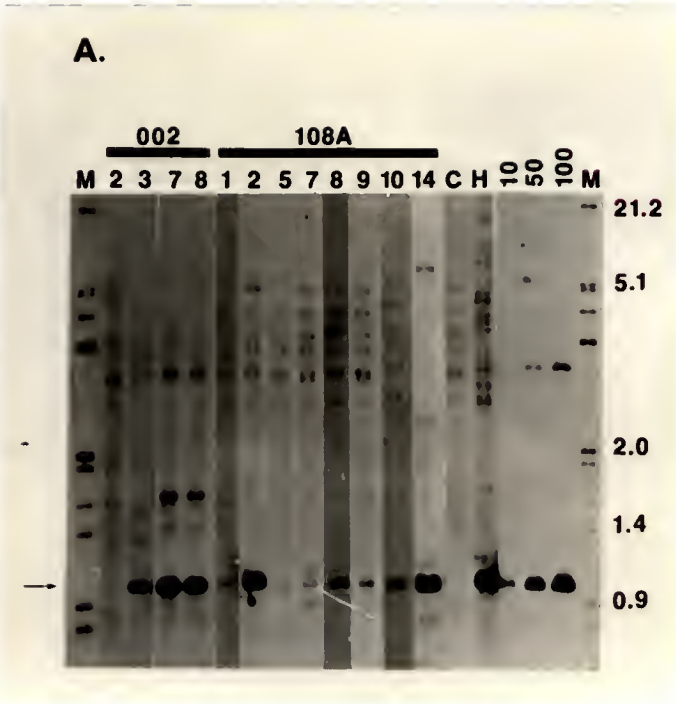


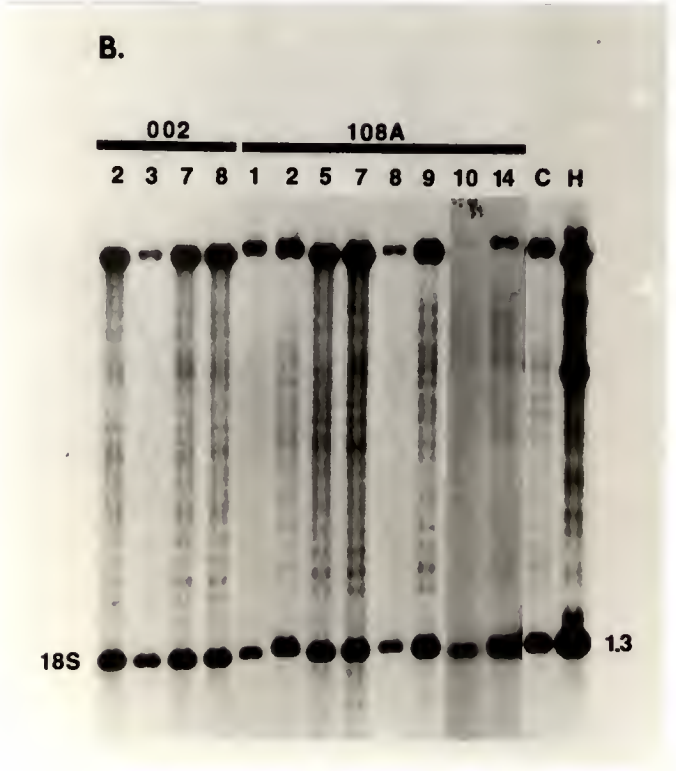
Figure 3-9 S1 nuclease analysis of pFO108A and pFO002 monoclonal cell lines.

S1 nuclease assays were performed as described in and Materials and Methods. The left panel is representative of results obtained from FO108A cell lines; the right panel with total cellular RNA from pFO002 cell lines. The human and mouse protected fragments are designated with labels and arrows. The markers, M, are pBR322 digested with HpaII and important sizes are noted. The number above each lane corresponds to the clone number of that construct. The markers were diluted M1:4 and M1:8 for densitometry quantitation purposes. H, HeLa total cellular RNA. C, C127 total cellular RNA.

Figure 3-10 Copy number analysis of pF0002 and pF0108A monoclonal cell lines.

Southern blot analysis was performed as described in Materials and Methods. 10 μg of DNA from each cell line were analyzed with nick translated EcoRI/XbaI fragment from pF0002. A. pF0108A and pF0002 cell line DNA probed with H4 sequences. B. The histone probe was removed and the blot was reprobed with the mouse 18S ribosomal fragment. Densitometry of the 1070 bp band specified by the arrow in A and the 18S ribosomal band in B permitted quantitation of the copy number through normalization to the amount of DNA actually loaded and transferred as described in the Materials and Methods. The figure in A is a composite of several exposures that reflects the actual copy number and accounts for original quantitation errors. The plasmid controls for quantitation are labelled 10, 50 and 100 designating the number of pg loaded. C, C127 cellular DNA. H, HeLa cellular DNA. M, λ DNA digested with EcoRI and Hind III and labelled with α-^{32}P-dCTP and Klenow fragment. Each set of clones is designated with the black bar and the number of the individual clones is above the lane.


A.
002
108A
M 2 3 7 8 1 2 5 7 8 9 10 14 C H 10 50 100 M
21.2
5.1
2.0
1.4
0.9


B.
002
108A
2 3 7 8 1 2 5 7 8 9 10 14 C H
18S
1.3

densitometry signals in Table 3-2 (p. 103). The average expression of nine pFO003 monoclonal cell lines, for which copy number was later determined, was 2.29 ± 2.43.

It was obvious that these results varied, so the copy number of each cell line was determined from the southern blots in Figure 3-8a,b. The Southern blots of pFO003 monoclonal cell line genomic DNA, digested with EcoRI and XbaI, were prepared as detailed earlier and in Materials and Methods. The hybridization probe was the 1070 bp EcoRI/XbaI fragment isolated from pFO002 and nick-translated. The actual copy number of each cell line was determined by densitometric analysis of the 1070 bp EcoRI/XbaI band with normalization for the amount of DNA actually loaded. The amount of DNA in each lane was determined by removal of the histone probe at 80°C in 0.1XSSC and subsequent hybridization with the oligo-labelled BamHI/SalI fragment of the mouse 18S ribosomal gene. Densitometry of the 18S ribosomal band (Figure 3-8b) permitted normalization of the histone H4 copy numbers and comparison to the plasmid controls for copy number (see Appendix A for sample calculation of copy number).

The copy number data helps to explain some of the variation seen with the original expression determination for each cell line. When pFO003 copy number is taken into account for the expression data in Table 3-2, the expression/copy ratio for all of the cell lines is lowered and the average expression/copy is 0.094 ± 0.091. It is apparent from the data in Table 3-2 that as copy number increases, the expression/copy increases until approximately 20-40 copies are present, after which it declines. The pFO003M15 cell line is perhaps lower than

expected with respect to expression because of an unusual or deleterious integration site. The threshold of expression at 20-40 copies indicated that a limited number of human histone genes could be integrated and expressed in any one cell. This phenomenon has been investigated further and is discussed later in light of genomic sequencing data presented in Chapter 4. Overall the pFO003 monoclonal cell lines had higher expression levels than other cell lines (compare expression values with others in Table 3-2), but the expression/copy was similar. Since copy number was implicated in the level of expression, we also calculated the average copy number of each group of monoclonal cell lines and this is presented in Table 3-2. The level of expression, as we have determined it here (Table 3-2), is a direct reflection of the copy number.

The results of the S1 analysis of the pFO108A and pFO002 monoclonal cell lines are presented in Figure 3-9. Both cell lines expressed at a relatively low level and the numerical data are presented in Table 3-2. The average level of expression/copy for pFO108A is .079 ± .061 and for pFO002 is 0.045 ± 0.053. The data collected for the pFO108A monoclones were previously divided into two groups. Originally, there was a construct, designated J40, that after sequencing of the deletion points was found to be identical to pFO108A. Therefore, these data were incorporated into the 108A data base. It is interesting to note that pFO108A and J40 were thought to have different lengths of 5′ sequence and yet their expression was shown to be almost identical. This separation of the original observations lends a measure of confidence to the analysis process that has been used in these studies.

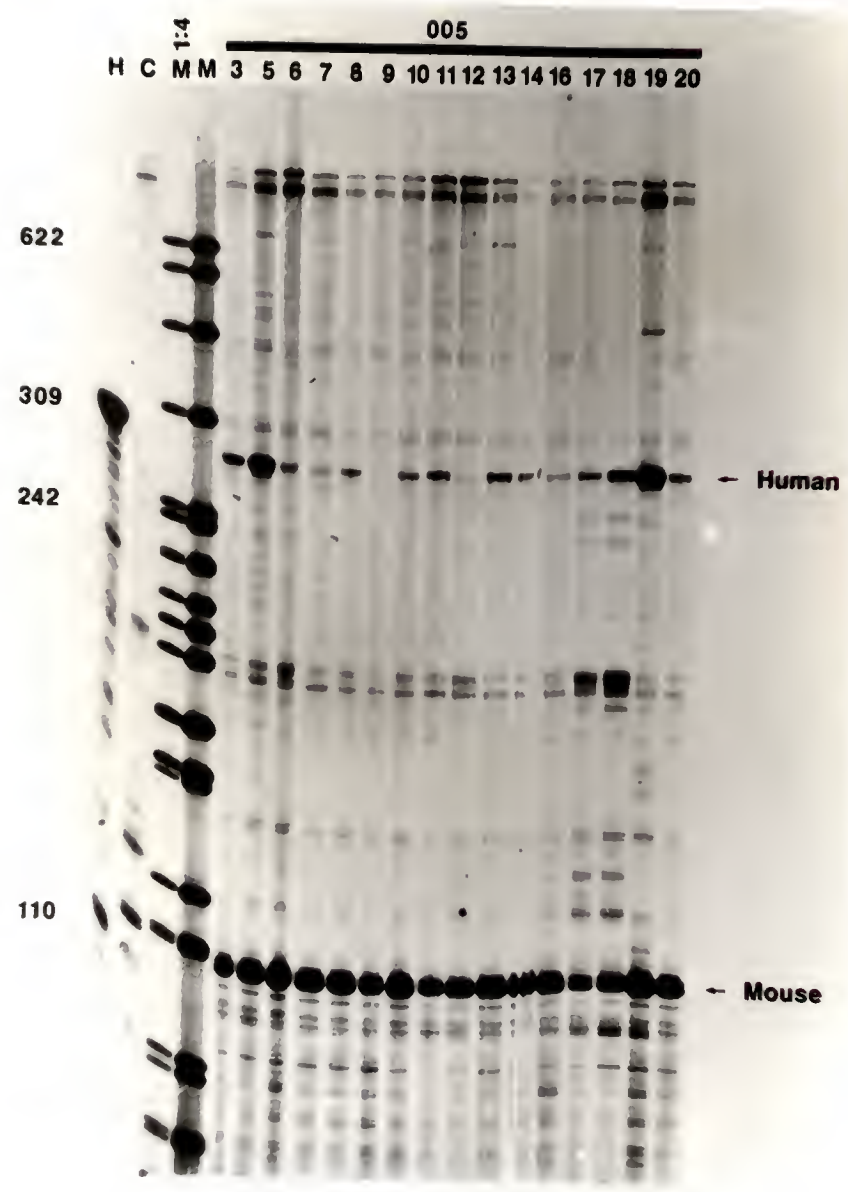
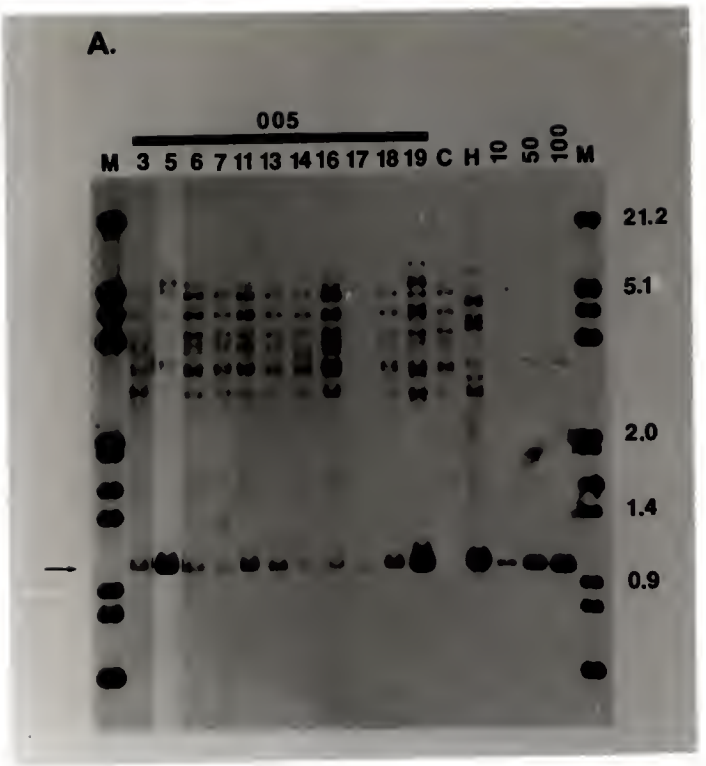


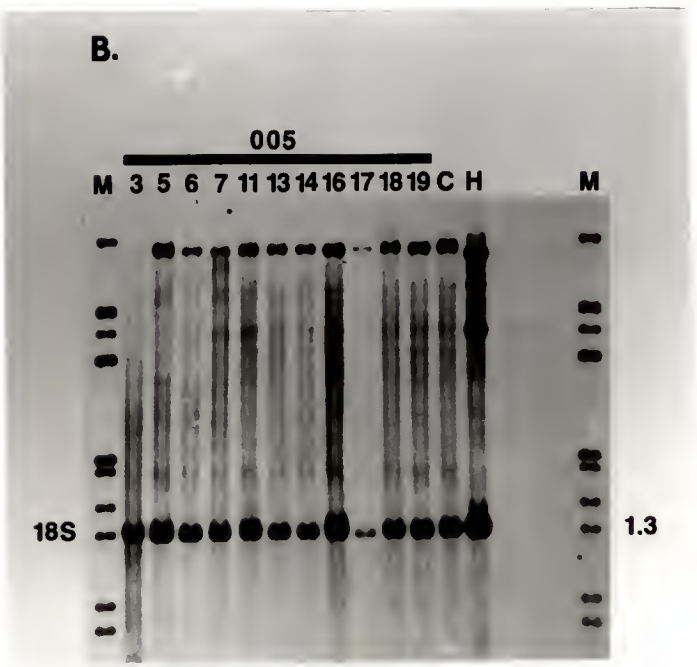
Figure 3-11 S1 nuclease analysis of pF0005 monoclonal cell lines.

Twenty five micrograms of total cellular RNA from each of the cell lines were treated as described in Materials and Methods and the autoradiograph of the S1 nuclease analysis was quantitated by densitometry. Lanes are designated with the clone number of the cell line. HeLa cell total RNA hybridized to both human and mouse probes, H. C127 RNA hybridized to both human and mouse probes, C. pBR322 HpaII markers labelled with α-^{32}P-dCTP and Klenow fragment, M. One fourth the amount of marker was electrophoresed for quantitation purposes, M1:4. The construct name, pF0005, is displayed above the black line. Both human and mouse (280 nt and 110 nt respectively) protected fragments are noted at the right.

Figure 3-12 Copy number analysis of pF0005 monoclonal cell lines.

Southern blot analysis was done as described in Materials and Methods. All abbreviations are as designated in Figure3-10. The quantitation of the histone H4 blot (A) and the mouse 18S ribosomal probed blot (B) are as before in Fig 3-10 and Materials and Methods.


A.
005
M 3 5 6 7 11 13 14 16 17 18 19 C H 10 50 100 M
21.2
5.1
2.0
1.4
0.9


B.
005
M 3 5 6 7 11 13 14 16 17 18 19 C H M
18S
1.3

The expression data for pFO002 monoclones 9 and 10 were determined (Table 3-2); however, when the copy number was determined there was no correct band at 1070 bp or any additional bands that corresponded to the EcoRl/XbaI fragment (data not shown). The copy number of the pFO002 and pFO108A cell lines (Figures 3-10a,b) was determined as described for pFO003 and in Materials and Methods. With respect to pFO002m9 and 10, we assume that either the construct was lost in the time between the harvesting of cells for the purification of RNA and subsequently DNA, or that the integration event destroyed one of the restriction sites making detection impossible. This was the only case where expression of the human H4 histone gene was detected but no copies were detectable. Due to the constraints of the tissue culture system we usually prepared several plates of cells for the isolation of RNA, and then 1 or 2 passages later was able to harvest cells for isolation of DNA. The lanes of the pFO002 Southern blot exhibited no other bands that might have corresponded to the integrated pFO002 construct.

S1 nuclease analysis and copy number determination were also done for the pFO005 monoclonal cell lines and the results are presented in Figures 3-11, and 3-12a,b. pFO005 exhibited the most consistency in the level of expression (0.546 ± 0.354) and nearly every monoclonal line was positive for expression of the human H4 gene. The expression/copy ratio was 0.201 ± 0.140.

The shortest deletion construct for which monoclonal cell lines were made was K8, an original Bal31 deletion (Sierra et al., 1983). The expression from all six monoclones measurable was relatively low

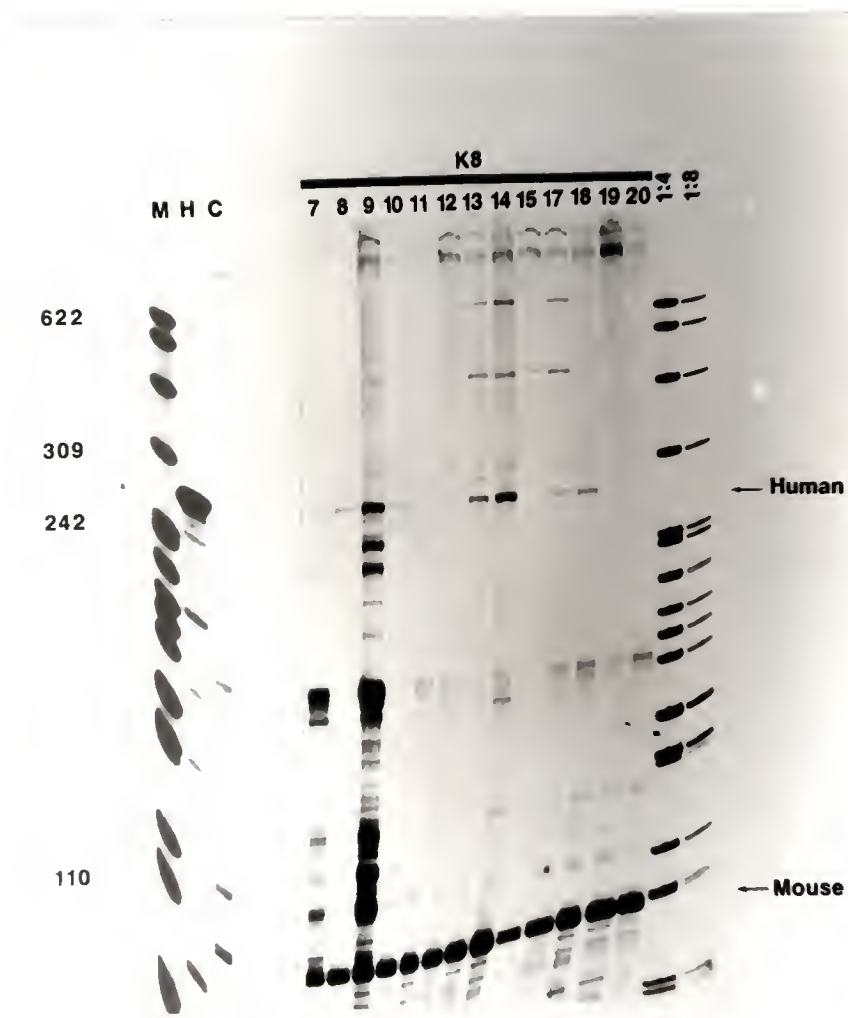
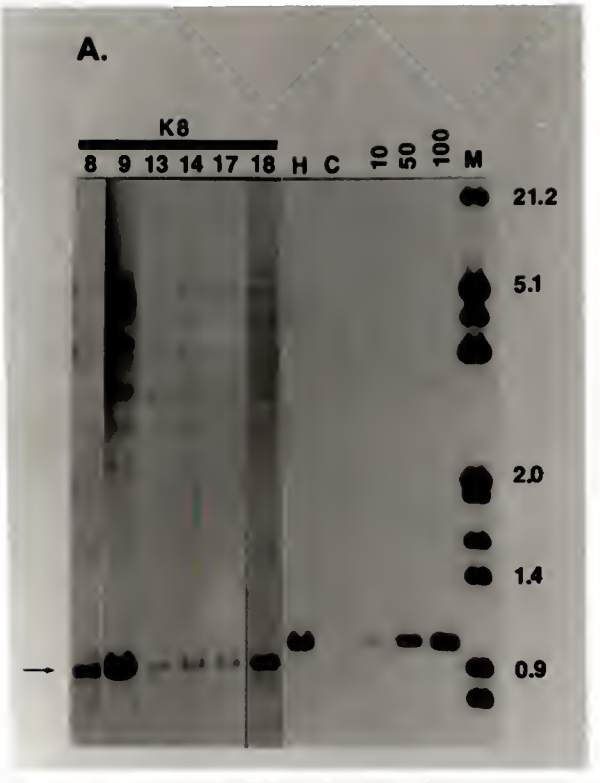


Figure 3-13 S1 nuclease analysis of K8 monoclonal cell lines.

S1 nuclease assays were done as described in Materials and Methods. The clone number of each cell line is denoted above the lane. M, pBR322 digested with HpaII and labelled with Klenow. H, HeLa total cellular RNA. C, C127 total cellular RNA. Dilutions of the marker are specified 1:4 and 1:8. Human and mouse H4 protected fragments are specified.

Figure 3-14 Copy number analysis of K8 monoclonal cell lines.

Southern blot analysis was performed as described in Materials and Methods. The K8 EcoRI/XbaI fragment is shorter due to the Bal 31 deletion and is designated with an arrow at the left. The same controls as in Figure 3-10 have been included. Quantitation of A (histone H4 probe) and B (reprobed with mouse 18S ribosomal) was as described in Figure 3-10 and Materials and Methods. Nonessential lanes in B have been deleted.


A.
K8
8 9 13 14 17 18 H C 10 50 100 M
21.2
5.1
2.0
1.4
0.9
B.
K8
8 9 13 14 17 18 H C
18S
1.3

(Table 3-2. Expression = 0.114 ± 0.066, Expression/Copy number = 0.075 ± 0.077). In addition, there were several K8 monoclones, including K8m12, 19, and 20, in which there was an S1 nuclease protected fragment present by visual inspection, but the level was below that detectable with the densitometer. The S1 nuclease protection assay and Southern blot analysis are presented in Figure 3-13 and 3-14a,b. These results were in agreement with the previous polyclonal cell line results that we had obtained that suggested that an increase in the length of the H4 promoter resulted in increased expression.

We were also concerned that differences in the 3′ end of some of our constructs might affect the level of expression. The differences in the 3′ ends of the constructs were not intentional, but arose as a result of the cloning strategies employed to produce the 5′ deletions. To address this question we prepared the construct pFO108X (see Appendix C). This construct has -210 bp of 5′ flanking sequence, but the XbaI/HindIII fragment at the 3′ end has been deleted from pFO108A. Also, this construct was made in pUC19. This 3′ deletion effectively removes 770 bp from the 3′ flanking region of the pFO108A H4 gene. Monoclonal cell lines of pFO108X were prepared and assayed for expression and copy number as before. The results of the analysis are presented in Figures 3-15 and 3-16a,b and the expression levels are calculated in Table 3-2. The expression of pFO108X was not significantly different than that of pFO108A ; these results suggest that the nucleotides from the XbaI (+1107 bp) site to the HindIII (+1877 bp) site, removed from the 3′ end in pFO108X, had little if any effect on the level of transcription. The construct pFO006 was also

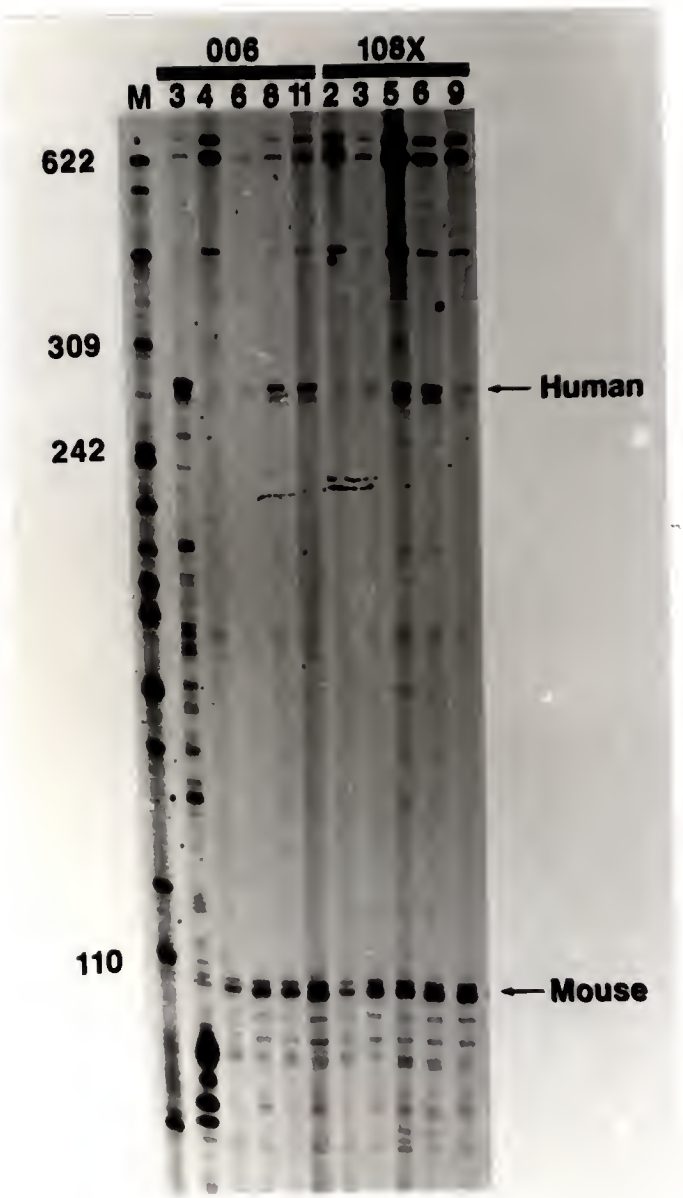


Figure 3-15 S1 nuclease analysis of pFO006 and pFO108X monoclonal cell lines.

S1 analysis of pFO006m3,4,6,8,11 and pFO108Xm2,3,5,6,9 are presented. 25μg of total cellular RNA from log phase cells was mixed with an excess of human and mouse H4 histone probes and S1 nuclease reactions electrophoresis, and densitometry were done as described in Materials and Methods. The human and mouse signals (280 nt and 110 nt respectively) are noted on the right. The markers, M, and pBR322 digested with HpaII and labelled with Klenow fragment and $^{\alpha}$-^{32}P-dCTP. The number above the lane designates the clone number, and the black line defines the construct.

Figure 3-16 Copy number analysis of pF0006 and pF0108X monoclonal cell lines.

Southern blot analysis was done as described in Materials and Methods. Quantitation of A and B (A, probed with human H4 histone; B, probed with mouse 18S ribosomal gene) was as described before in Figure 3-10. All designations are as described earlier in Figure 3-10.

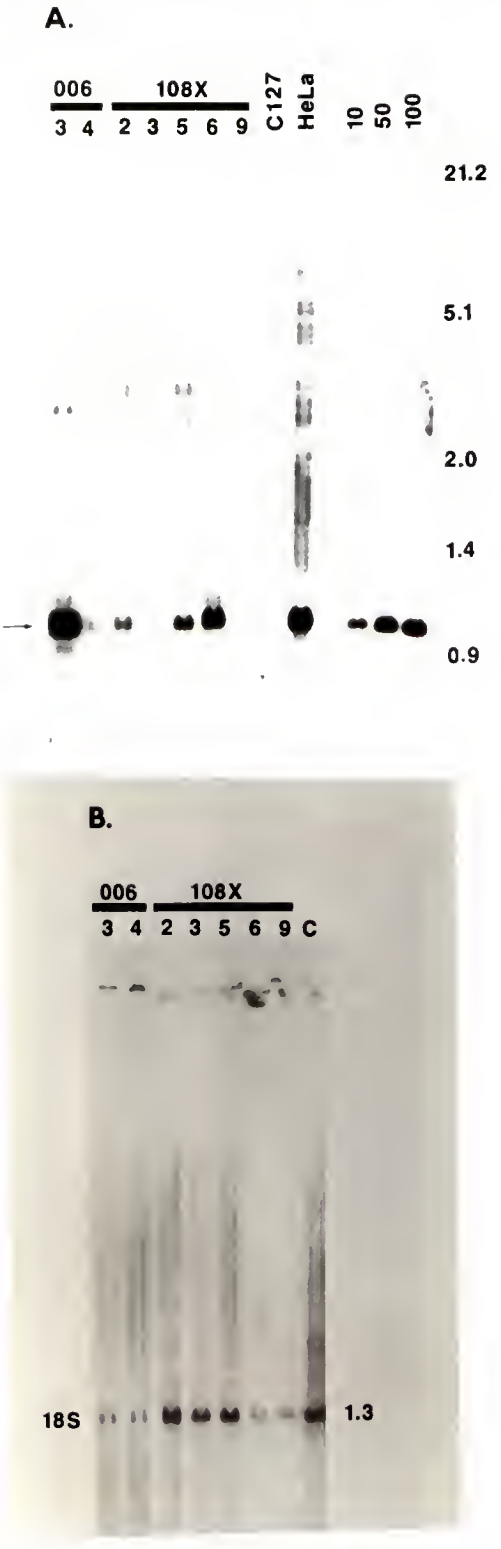

A.
006
108X
3 4 2 3 5 6 9
C127
HeLa
10
50
100
21.2
5.1
2.0
1.4
0.9
B.
006
108X
3 4 2 3 5 6 9 C
18S
1.3

Table 3-2. Quantitation of Monoclonal Cell Line Expression

CLONE[a]	EXP[b]	CN[c]	EXP/CN[d]	CLONE	EXP	CN	EXP/CN
K8m13	.091	1	.091	002m9	.036		ND
K8m17	.180	1	.180	002m10	.135		ND
K8m14	.170	1	.170	002m2	.135	1	.135
K8m18	.030	5	.006	002m3	.175	7	.025
K8m8	.032	13	.002	002m8	.179	14	.013
K8m9	.180	28	.002	002m7	.084	16	.005
AVG[e]	.114	8	.075	AVG	.124	10	.045
STD[f]	.066		.077	STD	.050		.053
108Am7	.123	1	.123	003m17	.100	4	.025
108Am1	.013	1	.013	003m13	.300	8	.038
108Am5	.056	1	.056	003m4	1.800	10	.180
108Am9	.210	1	.210	003m5	.500	13	.038
108Am10	.110	1	.110	003m14	1.770	21	.084
108Am12	.143	4	.036	003m16	6.780	23	.291
108Am8	.410	4	.103	003m2	6.500	41	.159
108Am2	.940	19	.049	003m15	.400	44	.009
108Am14	.240	30	.008	003m1	2.500	139	.018
AVG	.249	7	.079	AVG	2.286	34	.094
STD	.268		.061	STD	2.433		.091
005m14	.450	1	.450	007m1	.234	ND	--
005m16	.310	1	.310	007m2	.208	ND	--
005m7	.120	1	.120	007m4	.142	ND	--
005m11	.380	2	.190	007m8	.034	ND	--
005m6	.160	2	.080	007m9	.135	ND	--
005m3	.780	2	.390	007m10	.123	ND	--
005m13	.970	3	.323	007m12	.957	ND	--
005m18	.810	5	.162				
005m17	.430	5	.086	AVG	.261		
005m19	.630	22	.029	STD	.313		
005m5	1.280	31	.041				
AVG	.546	6	.201				
STD	.354		.140				

Table 3-2 continued

CLONE	EXP	CN	EXP/CN
108Xm2	.086	1	.086
108Xm3	.021	1	.021
108Xm9	.021	1	.021
108Xm5	.226	2	.113
108Xm6	.208	30	.006
AVG	.112	7	.049
STD	.089		.042

CLONE	EXP	CN	EXP/CN
004m6	.075	1	.075
004m14	.250	1	.250
004m2	.050	11	.005
004m8	2.38	38	.063
004m11	.125	90	.001
004m10	.175	154	.001
004m19	5.4	188	.029
004m1	1.27	252	.005
AVG	1.22	92	.054
STD	1.76		.079

006m3	5.050	30	.168
006m6	.013	ND	--
006m8	.158	ND	--
006m11	.102	ND	--
AVG[g]	.091		

004Rm3	.020	ND	--
004Rm4	.040	ND	--
004Rm7	.110	ND	--
004Rm9	.045	ND	--
004Rm10	.040	ND	--
AVG	.057		
STD	.031		

Autoradiograms were scanned with a laser densitometer as described in Chapter 2 and the level of expression has been determined and presented here as a ratio of the human and mouse S1 signals.

a. The construct used to make the cell line and the clone number that designates that cell line.
b. Expression (EXP): a ratio of the human and mouse S1 protected fragments as determined by densitometry. Further description of the calculation and densitometry procedures are given in the Materials and Methods section.
c. The copy number (CN) of the cell line as determined by Southern blot analysis.
d. Expression divided by the copy number of the cell line (EXP/CN). The number represents the level of expression per copy of the construct integrated.
e. AVG, the average of either EXP or EXP/CN.
f. STD, the standard deviation of the average value.
g. The data from 006m3 were not included in the average calculation.

assayed with the pFO108X, but will be discussed later in the chapter with respect to a putative enhancer element.

The data collected in Table 3-2 were arranged and the constructs placed into a rank order in comparison with one another. This is graphically presented in Figure 3-17. Both the average level of expression and expression/copy are presented with the standard deviation of each calculation. The first observation is that as the amount of 5' sequence is extended out to -410 bp in construct pFO005 the level of expression increases approximately 3 fold above that of pFO108A. There are significant differences between expression and expression/copy. This is most easily seen when copy number is included in the expression value for pFO003. There is obviously a large standard deviation in these results and this is probably a reflection of the inaccuracies inherent in the system available.

There are statistical differences in spite of the high variability encountered in this assay system. To analyze the data statistically we employed the services of the University of Florida Biostatistics Unit and Dr. Mike Conlon. It was decided that the most powerful statistical test that could be employed on these data was the Wilcoxon Rank Sum test. This test and the analyses have been described in the Materials and Methods section of this work.

Previously we demonstrated that the K8 monoclones (-155 bp) and the pFO108A monoclones (-215 bp) were not significantly different; these results suggested that there was little contribution from the sequence between -155 bp and -215 bp to the level of transcription. When the data for the pFO005 monoclones were compared to the K8 monoclones, the

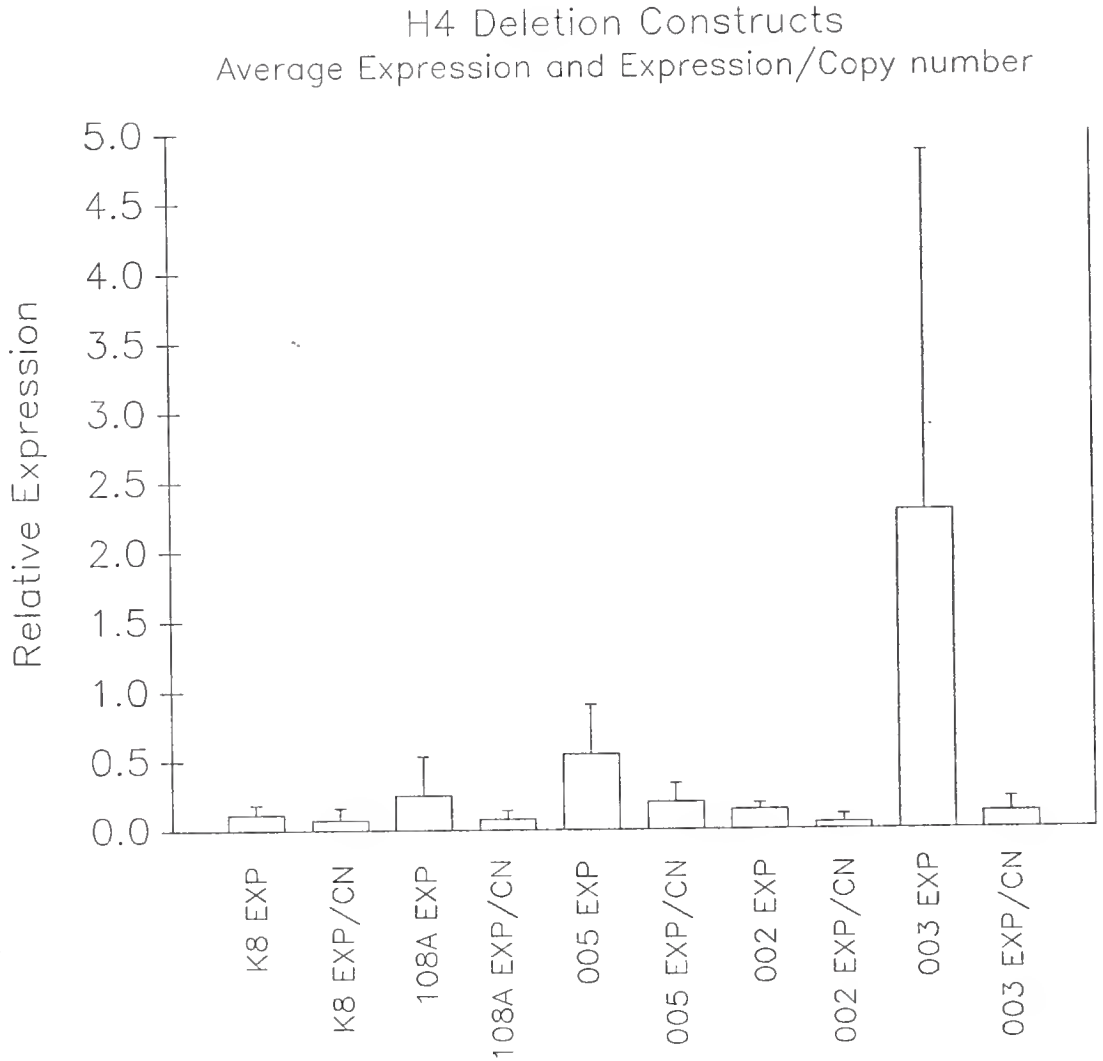


Figure 3-17 Graphic analysis of human H4 histone gene expression in mouse C127 cells.

The average expression for each group of monoclones is plotted. The average expression (EXP) and expression/copy (EXP/CN) were calculated in Table 2 and are plotted here with the standard deviation for each average value shown as a one-way error bar.

difference in the average expression/copy (pFO005 = 0.201 versus K8 = 0.075) was significant at $p < 0.05$. This suggested the existence of a positive regulatory element in the sequences from -215 bp to -417 bp.

Previous polyclonal cell line analysis had demonstrated a difference between pFO005 and the shorter deletions however the number of samples was small and precluded any statistical analysis. Since copy number had been demonstrated to be an important variable, we compared the pFO108A and pFO005 monoclonal cell lines with copy numbers less than 10 in the Wilcoxon Rank Sum test, and found that in this group of data, pFO005 was significantly ($p < 0.05$) higher in expression than pFO108A. If the entire data base was utilized, then the two constructs were not significantly different ($p < 0.1$). This was presumably the result of several high copy number cell lines that skewed the group. The data were consistent with the idea that a gradual increase in the 5' flanking sequences contributes to an increase in the level of human H4 histone gene expression. The data also suggested that there might be a positive regulatory element between -210 bp (pFO108A) and -410 bp (pFO005) although it was not clearly definable.

Protein/DNA interactions in this region of the promoter were detected *in vitro* by van Wijnen et al. (1987), unfortunately these studies were not pursued. *In vivo*, there were no detectable protein/DNA interactions in the -210 bp to -410 bp region (Pauli et al., 1987) of the promoter. We have done preliminary investigation into the putative positive element and our studies are detailed below. In addition, the data presented in Table 3-2 were reevaluated with respect to the effect of copy number on expression; the low copy

number data, which were most representative of the results and least affected by the competition phenomenon mentioned earlier, are discussed and analyzed in chapter 5.

Distal Transcriptional Regulatory Elements

Inspection of the data in Table 3-2 and graphically presented in Figure 3-17 demonstrates two points. First, as the length of the H4 promoter sequence increases to -410 bp the average level of expression rises. The difference in expression between the monoclonal cell lines of pFO108A and pFO005 is statistically significant ($p < 0.05$). The second point is that the expression of pFO002 is significantly lower than pFO005. This result is based on the comparison of only 4 of the 6 monoclonal cell lines which were positive for expression. Unfortunately, two of the pFO002 monoclones (pFO002m9 and pFO002m10) had no detectable EcoRI/XbaI fragment in the copy number experiment. We should note that if one assumes that these cell lines had only 1 copy of the H4 gene integrated and incorporates all 6 monoclonal cell lines into the data base, the difference in expression between pFO005 and pFO002 is still statistically significant. The fact that pFO002 (-1065 bp) was lower in expression than pFO005 (-417 bp) suggested that there might be a negative regulatory element in the more distal sequences of the H4 promoter.

The objective of the next experiments was to determine if there was a negative regulatory element located between -410 and -1065 bp in the human H4 histone gene promoter. There were several other lines of evidence that suggested that sequences upstream of -410 bp might influence the expression of this gene in a negative fashion. When we

Figure 3-18 Strategy utilized to determine the sequence of the BamHI (-1065 bp)/EcoRI (-610) fragment upstream of the FO108 H4 histone gene.

The pertinent restriction enzyme sites are designated. The arrows indicate that the DNA was restricted at the origin of the arrow and sequenced from that point in the direction of the arrow as described in Materials and Methods.

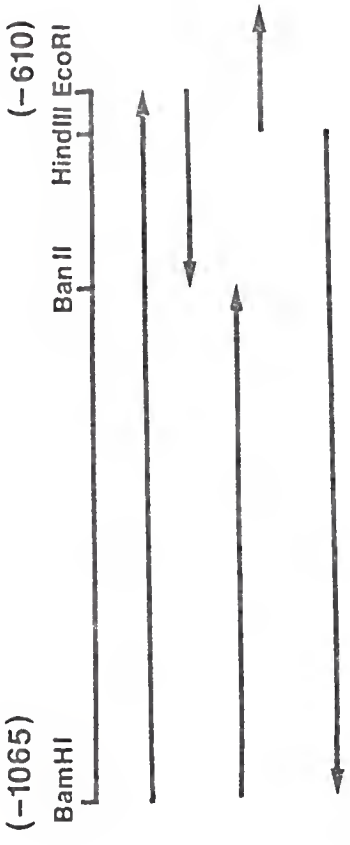

(-1065)
BamHI
Ban II
HindIII EcoRI
(-610)

Figure 3-19 Annotated sequence of the pFO002 5′ flanking sequence.

Both strands of the sequence are shown from +70 bp to -1010 bp. Construct deletion points are denoted with an asterisk over the last base in the clone and the clone name. A number of homologies to various elements have been designated. The proximal 210 bp have the ATG, TATA, CAAT, and GGTCC elements underlined. Also Site I and Site II are bolded. From -720 to -820 bp a DNaseI hypersensitive site is denoted with a string of asterisks above the sequence. Putative nuclear matrix attachment sites, underlined, are located at -680 bp and -940 bp (bases that do not match the consensus sequence are bolded). A putative topoisomerse II site is found at -881 bp to -895 bp and has been confirmed by Dr. T. Rowe (personal communication). Nuclear matrix associated T-boxes are underlined at -925 and -885 bp. Two putative negative regulatory elements are underlined at -580 and -710 bp.

```
*002 (-55bp)
GCAGAATATCCCTCAGTCTTCTCTATGTAGCAGGCCCTCCATATACGCGGGTTCCCCAAG
  .          -1000      .          .          .          .
CGTCTTATAGGGAGTCAGAAGAGATACATCGTCCGGGAGGTATATGCGCCCAAGGGGTTC

                                        *002D1
ACCGAAAATATTAAACAAATGAATTTCTTTTTTAAAAAAAAGTACAACAAAAGATAGTAA
 -950       .          .          .          .          -900
TGGCTTTTATAATTTGTTTACTTAAAGAAAAAATTTTTTTTCATGTTGTTTTCTATCATT

AAATAAAAACAGTATAACAATTACTTACATAGCTTTACACACTGGATTGGTGTTCGAAGT
  .          .          .          .          -850       .
TTTATTTTTGTCATATTGTTAATGAATGTATCGAAATGTGTGACCTAACCACAAGCTTCA

           **********************************************************
AATTTGAGCTTATTTAAAGTACACGGGAGGATGTGCATAGTTATGTGCAAATACTACCCC
  .          .          .          -800       .          .
TTAAACTCGAATAAATTTCATGTGCCCTCCTACACGTATCAATACACGTTTATGATGGGG
                                                *002E9
**********************************************************************
ACTTTCTATGAGAGACTTGAGCAACCTGATTTTGGTATCGGCGGGGGCCCTGACCAATCC
  .          .          -750       .          .          .
TGAAAGATACTCTCTGAACTCGTTGGACTAAAACCATAGCCGCCCCCGGGACTGGTTAGG

CCTCTCAGTTCTACCGAGGGAGAACTGTTTTGTTTCTTCCGCACGGCTTTGACCGACAGT
  .          -700       .          .          .          .
GGAGAGTCAAGATGGCTCCCTCTTGACAAAACAAAGAAGGCGTGCCGAAACTGGCTGTCA

GTGTTGGGATTCGCTGGACCATGAGAAAGCTTGGCAGCATGCTGTGACCGGTTTTCCCAG
 -650       .          .          .          .          -600
CACAACCCTAAGCGACCTGGTACTCTTTCGAACCGTCGTACGACACTGGCCAAAAGGGTC

     *007
GGCCAGAATTCTCCTGTGTGAGCTAAAATACAGTGGCTCGGTCCAACAAAACAGAGCCTG
  .          .          .          .          -550       .
CCGGTCTTAAGAGGACACACTCGATTTTATGTCACCGAGCCAGGTTGTTTTGTCTCGGAC

GAGCCAGGAATTATGGCGAACCTGCTCCCTCCGTCCTCCTTCGGCGAAGATCCCTGGCGC
  .          .          .          -500       .          .
CTCGGTCCTTAATACCGCTTGGACGAGGGAGGCAGGAGGAAGCCGCTTCTAGGGACCGCG

                                                          *005
GCGTCCTTGAGGTCGCCTTCGGTGTTGACCTCATCGTCGGAACGGCGCTTCCTGAAGCTT
  .          .          -450       .          .          .
CGCAGGAACTCCAGCGGAAGCCACAACTGGAGTAGCAGCCTTGCCGCGAAGGACTTCGAA
```

```
TATATAAGCACGGCTCTGAATCCGCTCGTCGGATTAAATCCTGCGCTGGCGTCGTGCCAG
  ·          -400       ·          ·          ·          ·
ATATATTCGTGCCGAGACTTAGGCGAGCAGCCTAATTTAGGACGCGACCGCAGGACGGTC

TCTCTCGCTCCATTTGCTCTTCCTGAGGCTCCCTCCAGAGACCTTTCCCTTAGCCTCAGT
 -350       ·          ·          ·          ·          -300
AGAGAGCGAGGTAAACGAGAAGGACTCCGAGGGAGGTCTCTGGAAAGGGAATCGGAGTCA

GCGAATGCTTCCGGGCGTCCTCAGAACCAGAGCACAGCCAAAGCCACTACAGAATCCGGA
  ·          ·          ·          ·          -250       ·
CGCTTACGAAGGCCCGCAGGAGTCTTGGTCTCGTGTCGGTTTCGGTGATGTCTTAGGCCT

                   *108A                                  *L14
AGCCCGGTTGGGATCTGAATTCTCCCGGGGACCGTTGCGTAGGCGTTAAAAAAAAAAAAG
  ·          ·          ·          -200       ·          ·
TCGGGCCAACCCTAGACTTAAGAGGGCCCCTGGCAACGCATCCGCAATTTTTTTTTTTTC

                   *K8
AGTGAGAGGGACCTGAGCAGAGTGGAGGAGGAGGGAGAGGAAAACAGAAAAGAAATGACG
  ·          ·          -150       ·          ·          ·
TCACTCTCCCTGGACTCGTCTCACCTCCTCCTCCCTCTCCTTTTGTCTTTTCTTTACTGC

            *J50                                  *J56
AAATGTCGAGAGGGCGGGGACAATTGAGAACGCTTCCCGCCGGCGCGCTTTCGGTTTTCA
  ·          -100       ·          ·          ·          ·
TTTACAGCTCTCCCGCCCCTGTTAACTCTTGCGAAGGGCGGCCGCGCGAAAGCCAAAAGT

    *J67
ATCTGGTCCGATACTCTTGTATATCAGGGGAAGACGGTGCTCGCCTTGACAGAAGCTGTC
 -50        ·          ·          ·          ·          +1
TAGACCAGGCTATGAGAACATATAGTCCCCTTCTGCCACGAGCGGAACTGTCTTCGACAG

TATCGGGCTCCAGCGGTCATGTCCGGCAGAGGAAAGGGCGGAAAAGGCTTAGGCAAAGGG
  ·          ·          ·          ·          +50        ·
ATAGCCCGAGGTCGCCAGTACAGGCCGTCTCCTTTCCCGCCTTTTCCGAATCCGTTTCCC
```

Figure 3-19 continued

tried to make cell lines with the construct pF0001 (an internal deletion from -210 to -610 bp, see Figure 3-1), we found very few cell lines that expressed the transfected human H4 gene. Only 1 of 10 polyclonal lines and 2 of 12 monoclonal lines were positive for expression, and these were barely detectable (data not shown). This result supported the idea that there was a positive regulatory element in the region between -210 and -417 bp and, since the expression was very low perhaps a negative element upstream of -586 bp. In addition, polyclonal lines of pF0002 appeared to have lower expression than polyclonal lines of pF0005 (data not shown).

To address the possibility of a negative regulatory element we first decided to sequence the region of the promoter from -610 bp to -1065 bp. pF0002 DNA was digested with either BamHI, EcoRI, or HindIII, treated with phosphatase and labelled as described in Materials and Methods and the protocol of Maxam and Gilbert (1980). Figure 3-18 schematically displays the strategy utilized to determine the sequence of the upstream region. The sequence has several unusual characteristics and is presented as part of the entire pF0002 sequence in Figure 3-19. The distal end of the fragment, from -800 bp to -960 bp is very A/T rich (70%) with several homopolymeric runs of each. In addition a search of the region revealed two sequences with strong similarity to nuclear matrix attachment sites (-940 bp and -680 bp) and associated T-boxes (-925 bp and -890 bp, bottom strand)(Gasser and Laemmli, 1987). Near the upstream matrix site a putative topoisomerase site (-890 bp) was identified. The presence of this topoisomerase site has been confirmed <u>in vitro</u> by Dr. Tom Rowe

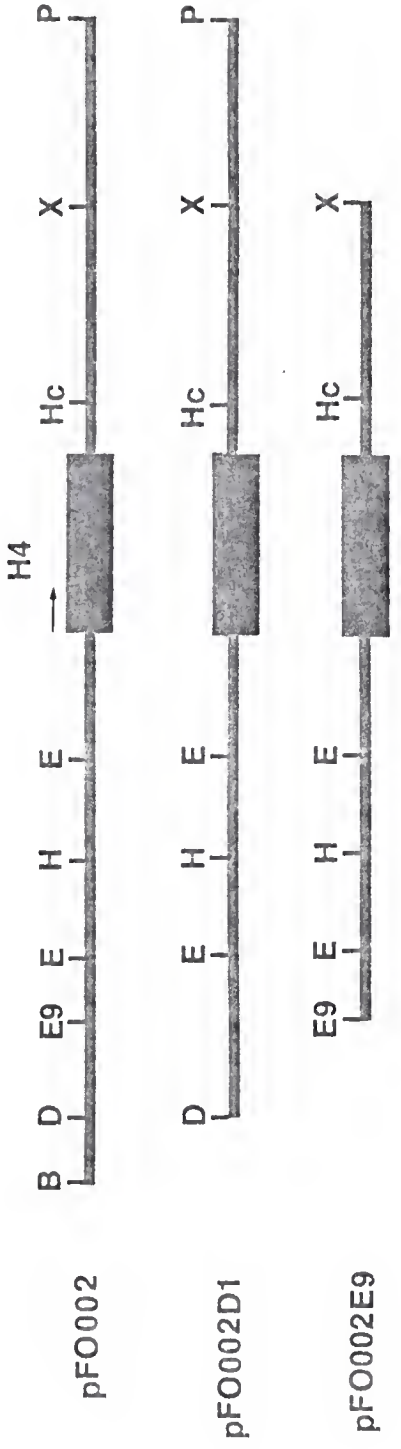
(personal communication). Additionally, Dr. Susan Chrysogelos of our laboratory demonstrated the presence of a DNAseI hypersensitive site (-720 to -820 bp) between the two putative nuclear matrix attachment sites (personal communication). This arrangement of chromatin structure, nuclear matrix sites flanking a nuclease hypersensitive site, is very similar to that demonstrated previously by Gasser and Laemmli (1986) and was at least circumstantial evidence that this region might be involved in attachment to the nuclear matrix.

We compared the entire 5' flanking sequence of the FO108 H4 histone gene (-1 to -1065 bp) with the consensus sequences for several groups of negative regulatory elements as described by Baniahmad et al. (1987). They compared the promoter sequences of a number of genes subject to negative regulation and determined two consensus elements. These elements were termed Box 1 (5'-ANCCTCTCC-3') and Box 2 (5'-ANTCTCCTCC-3'). Good homologies to both elements were found in the H4 histone upstream region at -710 bp (Box 1) and -580 bp (Box 2) as designated in Figure 3-19. Dr. Susan Chrysogelos, of our laboratory, demonstrated that the region of the H4 promoter from -585 to -1065 bp has middle repetitive character (personal communication). As mentioned in the introduction Laimins et al. (1986) associated middle repetitive character with some negative regulatory elements (chicken lysozyme gene and rat insulin like growth gene).

To investigate whether there was any functionality associated with the two putative negative regulatory elements that were implicated via similarity to previously identified negative elements we constructed two deletion mutants of pFO002 in the 460 bp BamHI/EcoRI fragment

Figure 3-20 Deletions for investigation of putative negative regulatory element in upstream region of H4 promoter.

Two deletions of pFO002 were made by standard cloning procedures as outlined in Materials and Methods and cloned into pUC19. The construct pFO002D1 is a DraI/PstI fragment that deletes 130 nucleotides from the 5' end of pFO002 in the region of interest. pFO002E9 is an Eco0109/XbaI fragment that deletes 315 nucleotides. The restriction enzyme sites are denoted as BamHI, B; DraI, D; Eco0109, E9; EcoR1, E; HincII, Hc; XbaI, X; and PstI, P. The direction of transcription is specified with the arrow.


B
D
E9
E
H
E
H4
Hc
X
P
pFO002
D
E
H
E
Hc
X
P
pFO002D1
E9
E
H
E
Hc
X
pFO002E9

(Figure 3-20). The first deletion, pFO002D1, was made by digestion of pFO002 with PstI and then a partial digestion with DraI. The DraI/PstI (2.16 kb) fragment was isolated and cloned into the SmaI site in pUC19. Before ligation the DNA was treated with Klenow fragment to blunt the ends of the insert molecule. This construct effectively deletes 145 bp of the 5' sequence from -1065 to -920 bp. The second deletion, pFO002E9, was prepared in a similar manner except the initial digestion was with XbaI. The partial digestion was (after the blunt end reaction with Klenow) with Eco0109. The 1630 bp Eco0109/PstI fragment was purified and the fragment was ligated to pUC19 digested with SmaI under blunt end conditions as described in Materials and Methods. pFO002E9 (-730) deletes 335 bp from the 5' end of pFO002. Just prior to the construction of pFO002D1 and E9 we made the construct pFO007 by an EcoRI partial digestion of pFO003 linearized with PstI. The 1.84 kb PstI/EcoRI fragment was cloned into pUC19 (Figure 3-1). This construct was made to assess the contribution of the 200 bp between -410 (pFO005) and -610. It was decided that instead of making stable cell lines, which had been a confusing endeavor up to this point, we would transiently transfect C127 cells with these constructs to assess any possible negative regulatory effects. The transfections of pFO002, pFO002D1, pFO002E9, pFO007, pFO005, pFO108A, and pFO001 were done according to the protocol described in Materials and Methods. Two of the S1 nuclease assays performed on RNA isolated from the transfected cells are presented in Figure 3-21. Both C127 cells and Ltk$^-$ cells were utilized in this series of experiments. The data from the series of transfections, 6 in total, are presented in Table 3-3. There was

Figure 3-21 S1 nuclease analysis of transiently transfected C127 and Ltk$^-$ cells: Determination of putative negative regulatory element position in distal promoter sequence of the FO108 H4 histone gene.

S1 nuclease analysis was performed on total cellular RNA of both C127 and Ltk$^-$ after transfection with 10 μg of each histone deletion construct as described in Materials and Methods. 50 μg of total cellular RNA was used for each hybridization reaction. The results of two transfections are presented. The human (280 nt) and mouse (110 nt) protected fragments are designated at the right. Markers (M), pBR322 digested with HpaII, labelled with α-^{32}P-dCTP and Klenow fragment, and pertinent sizes are shown. Densitometry of the human and mouse signals from autoradiograms permitted the quantitation of expression from each construct. The clones transfected into either C127 cells (left panel) or Ltk$^-$ cells (right panel) are specified above each lane. These two autoradiograms are representative of the 6 experiments that were performed. An analysis of the data is presented in Table 3 and Figure 3-22.

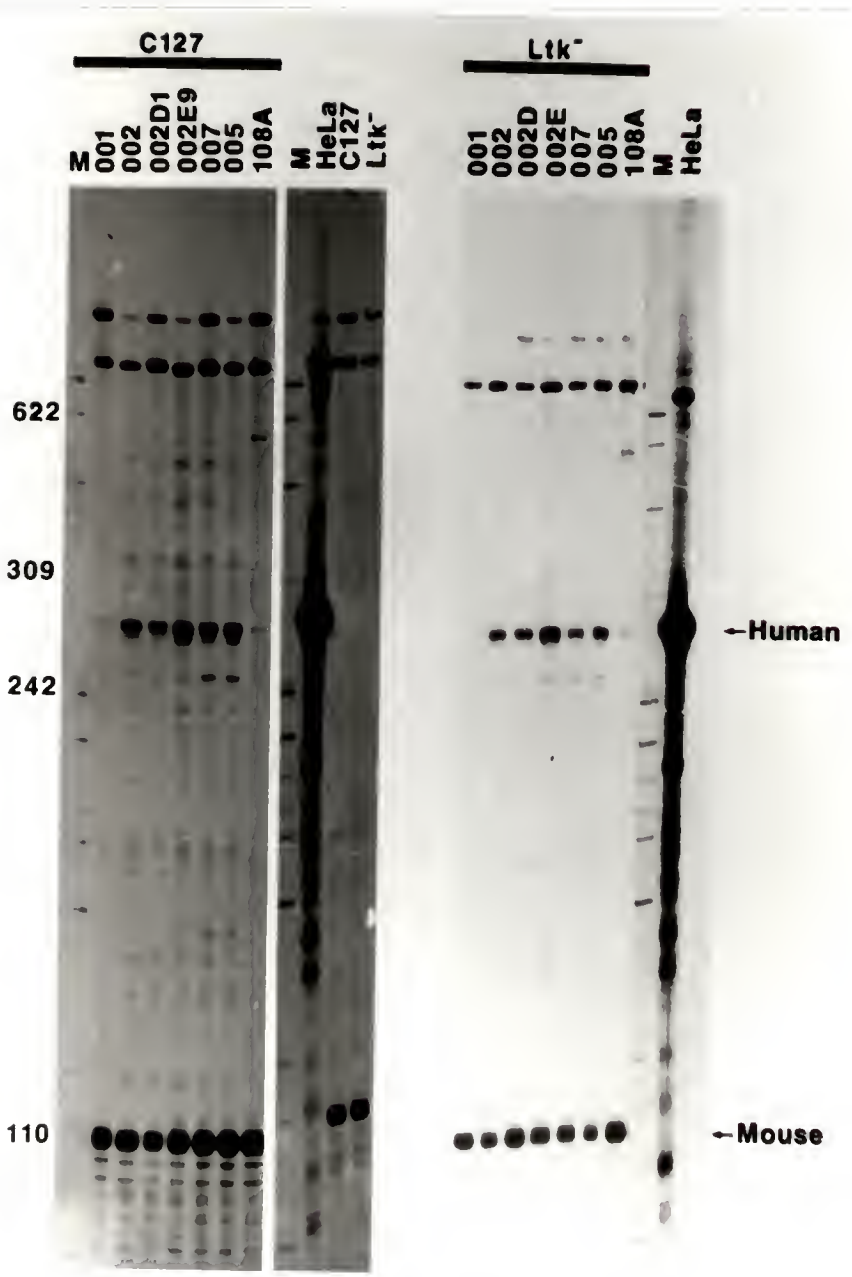

C127
Ltk⁻
M
001
002
002D1
002E9
007
005
108A
M
HeLa
C127
Ltk⁻
001
002
002D
002E
007
005
108A
M
HeLa
622
309
242
110
←Human
←Mouse

Table 3-3 Summary of Transient Expression Data.

	Experiment						
	1	2	3	4	5	6	
Cell Type	C127	Ltk$^-$	Ltk$^-$	C127	Ltk$^-$	C127	
Construct							Avg ± SD
pFO002	56	57	66	54	26	54	52.1 ± 13.5
pFO002D1	12	25	19	53	34	48	31.8 ± 16.2
pFO002E9	100	100	100	100	100	100	100.0
pFO007	68	34	14	42	--	56	42.8 ± 20.7
pFO005	9	72	42	49	23	52	41.1 ± 22.3
pFO001	ND	0.5	1	0.5	ND	ND	0.66 ± 0.3
pFO108A	ND	2	4	2	ND	ND	2.6 ± 1.1

Each construct was transfected into both C127 or Ltk$^-$ cells and analyzed by densitometry of the autoradiograms. The amount of transcription is expressed in percent of pFO002E9 expression since it was consistently the highest expressed construct. The individual values for each experiment are listed and the average expression (Avg) for the construct ± the standard deviation (SD) is listed at the right. Only 5 transfections of pFO007 were done; therefore there is no value for experiment #5. ND, not determined. Densitometry was only done on three of the experiments in which pFO001 and pFO108A were included.

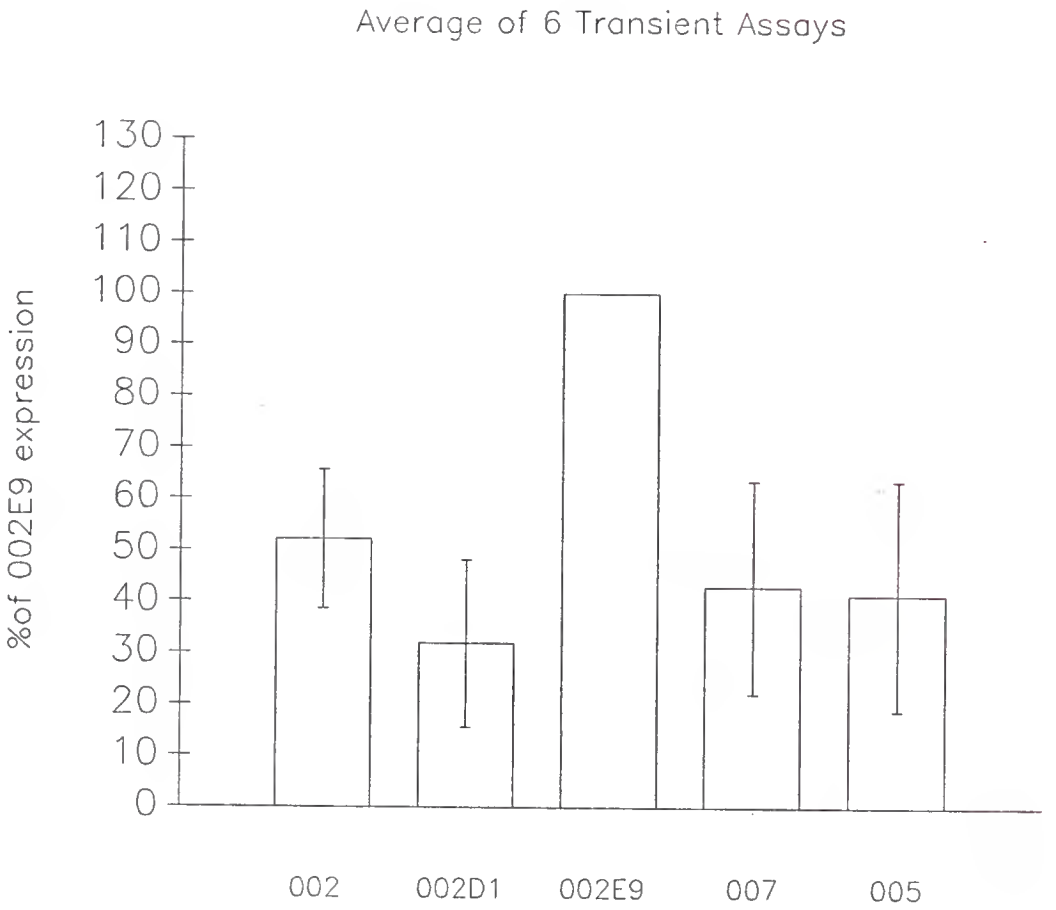


Figure 3-22 Compilation analysis of 6 transient assays with histone H4 deletion constructs: analysis of putative negative regulatory element.

The data from all six transient transfection experiments was averaged for each construct and plotted with standard deviation bars. The data, as in Table 3-3, are calculated as the percent of pF0002E9 expression. pF0002, pF0002D1, pF0002E9, pF0007, and pF0005 are included.

variability from experiment to experiment and the average of 6 experiments is plotted graphically in Figure 3-22. Plasmid DNAs were examined on agarose gels to determine the percent of Form I and to ensure that the quantitation was accurate. In both Table 3-3 and Figure 3-22 the values for pFO002, pFO002D1, pFO007 and pFO005 are expressed as the percentage of pFO002E9, which was consistently highest throughout the 6 experiments. The data for pFO001 and pFO108A were not included in Figure 3-22 as they were considerably lower than any of the other constructs; however the data are presented in Table 3-3.

Our first observation was that the level of pFO005 expression was very similar to that of pFO002 in apparent contrast to the data from the stable cell lines (Table 3-2) where pFO005 (-417 bp) had a 3 fold higher level of expression than pFO002 (-1065 bp). The most likely explanation for this appears to relate to differences between expression from stably integrated and episomal DNA molecules. This difference in the state of the DNA may also have affected the level of expression from pFO001 and pFO108A (discussed below). The original hypothesis was that the consensus negative regulatory sequences described earlier, Box 1 and Box 2 (Baniahmad et al., 1987), were responsible for the decrease in expression of the pFO002 construct. Our results demonstrated that both of the consensus negative regulatory elements (Box 1, -710 bp and Box 2, -580 bp) were located in the sequences included in the construct pFO002E9 and it was the most highly expressed construct of the group. This result disproves the idea that the decrease in expression is due to the proposed negative regulatory sequences. However, when additional sequences are added in pFO002D1

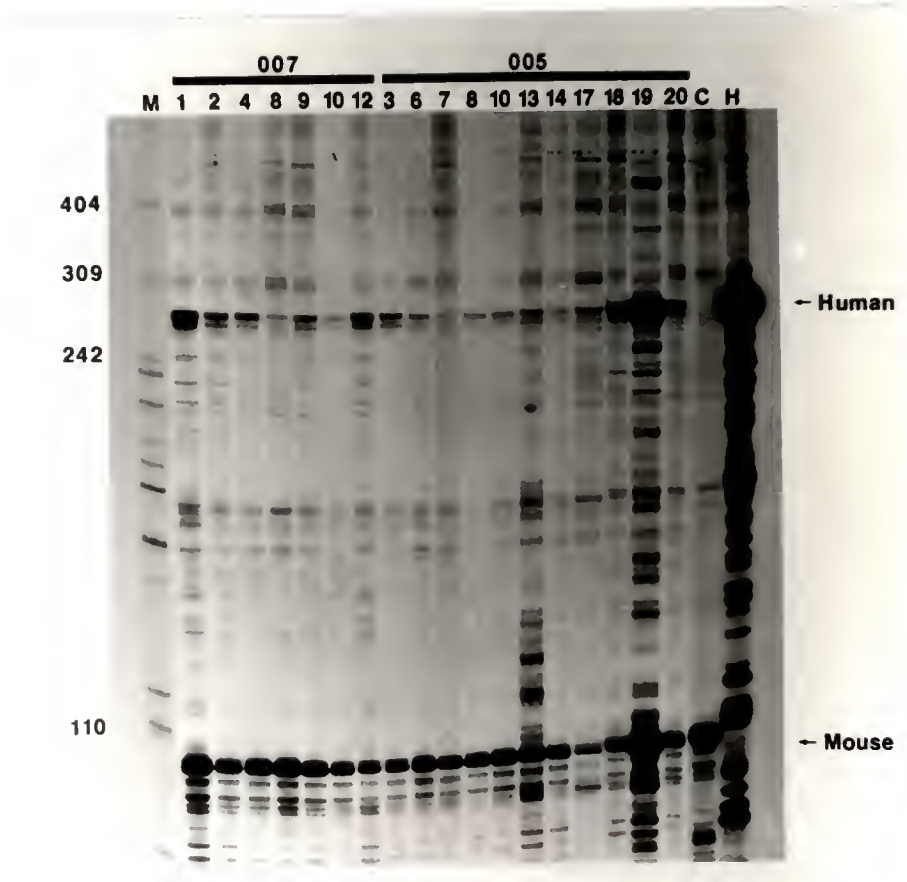


Figure 3-23 S1 nuclease analysis: Comparison of pFO007 and pFO005 monoclonal cell lines.

Analysis of total cellular RNA from each cell line was as described in Materials and Methods. Lanes are labelled as in previous S1 analysis figures: M, pBR322 HpaII marker; pFO007 and pFO005 monoclonal cell line numbers are denoted above the lane; H, HeLa RNA; C, C127 RNA. Both human and mouse S1 nuclease protected fragments are denoted to the right.

and pF0002 the expression is lower. If there is a negative element, and this evidence is again only suggestive for one, it probably lies in the sequences between the DraI site (-920) and the Eco0109 site (-730). Interestingly both the topoisomerase II site (-890 bp) and the DNaseI hypersensitive site (-720 to -820 bp) are included in this region of the promoter.

In addition, we can state that the sequences included in the construct pF0007 (-410 to -610 bp) do not contribute to the level of expression. In the transient assays the expression from pF0005 and pF0007 was nearly identical (Figure 3-22 and Table 3-3). The pF0007 construct was also transfected into C127 cells and monoclonal cell lines prepared. We compared the level of expression of the two constructs and found no significant difference. The S1 nuclease analysis is presented in Figure 3-23 and the expression data were calculated as before and displayed in Table 3-2.

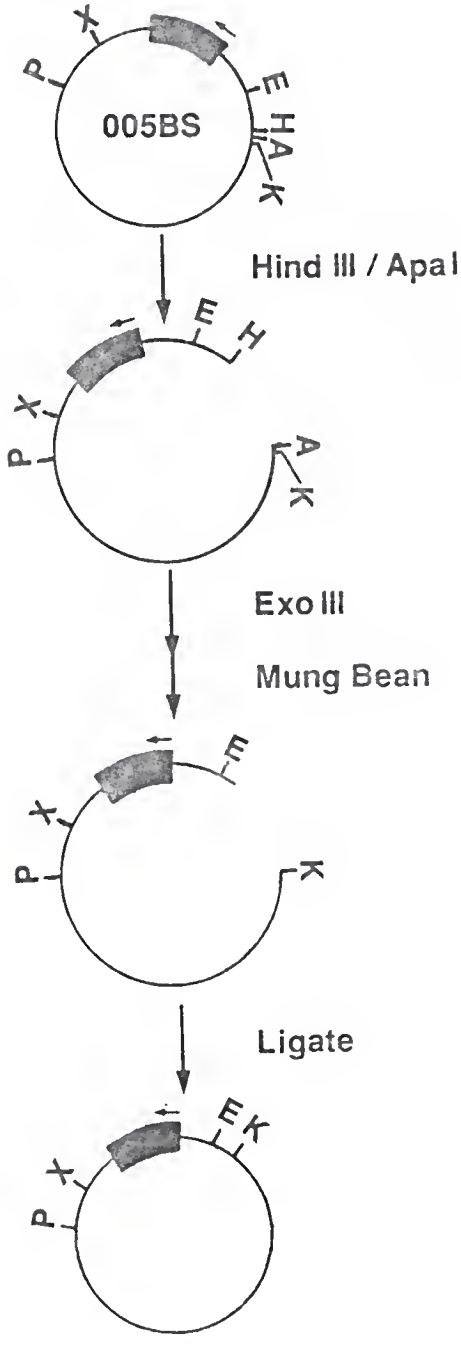
The construct pF0002E9 was consistently expressed at a higher level than any of the other constructs. An examination of the additional sequence included in the construct pF0002E9 revealed a putative CCAAT box (-700 bp) which matches the consensus sequence identically. Perhaps this element is responsible for the 2 fold increase in the level of expression. How the CCAAT box, normally a proximal promoter element, might function in this particular position is unknown; however there is precedence for the action of distal regulatory elements through bending of the DNA molecule (Ptashne, 1986). The results we have presented here are preliminary, but similar results have been shown by Ken Wright of our laboratory with <u>in vitro</u> transcription of

the same deletion constructs from circular templates in nuclear extracts (personal communication).

In the stable cell lines, the expression of a particular construct was apparently determined by the histone 5′ sequences and the number of copies integrated. In the transient assays, there was a depression in expression of pFO001 and pFO108A (Table 3-3) as compared to pFO005. This was consistent with the results of the stable cell lines, however the effect was more dramatic in the episomal system. pFO001 expression was low and usually imperceptible in stable cell lines, and consistent with this result was expressed only at a low level in the transient transfections. Even with the additional 5′ sequence that pFO001 includes the internal deletion of the EcoRI (-210)/EcoRI (-610) fragment again suggested that there was a positive element in the 200 bp between -210 and -410 nt. An alternative explanation for these results however is possible. pFO001 and pFO108A are different from the other deletion constructs in that they are both pBR322 clones whereas the others are all derivatives of pUC plasmids. Perhaps this difference in the length and composition of the vector was more dramatically accentuated in the transient assay system. In stable cell lines, a comparison of the 3′ deletions pFO108A and pFO108X revealed no significant differences ($p < 0.1$) in the level of expression (Table 3-2). If anything, pFO108X was slightly lower in expression and is a pUC19 clone. Perhaps the functionality of these upstream elements relies on a particular chromatin structure of the region which is only obtained when the constructs are integrated.

Figure 3-24 Schematic diagram for the production of unidirectional deletions with Exonuclease III and Mung Bean Nuclease.

The original construct pFO005B5 was made by insertion of the pFO005 insert into the PstI and HindIII sites of Bluescript M13+. This construct was digested with ApaI which produces a 3' overhang and HindIII, a 5' overhang as shown. Next 30 units of Exonuclease III were added and aliquots removed every minute for 5 minutes. The DNA aliquots were diluted in Mung Bean nuclease buffer and 9 units of Mung bean nuclease was added. The DNA was then ligated under blunt end conditions and transfected into competent DH5 cells. The complete protocol is detailed in Materials and Methods. The resulting products of this reaction are unidirectional deletions, because Exonuclease III is unable to digest a 3'single strand overhang (ApaI). Restriction enzyme sites are designated as PstI, P; EcoRI, E; HindIII, H; ApaI, A; and KpnI, K.


005BS
P
X
E
HA
K
Hind III / Apa I
E
H
X
P
A
K
Exo III
Mung Bean
E
X
P
K
Ligate
E
K
X
P

Distal-proximal positive element

Our results indicated that the sequences from -210 to -410 bp might enhance the level of H4 gene expression. To address this question deletion mutants of pF0005 were prepared. The pF0005 insert was subcloned into the Bluescript M13+ vector and deletions with Exonuclease III were prepared as described in Materials and Methods. The procedure is schematically displayed in Figure 3-24. Ideally, the protocol should have produced unidirectional deletions from the HindIII site at -410 bp toward the EcoRI site at -210 bp. This method relied on the inability of Exonuclease III to digest a 3′ overhang (ApaI). However the protocol worked very poorly, and only 2 deletions were obtained in the region of interest (-210 to -410 nt). These were denoted pF0005BSdel2-6 (-285 bp) and pF0005BSdel2-10 (-335 bp). Monoclonal cell lines of pF0005BS, pF0005BSdel2-10 (2-10), and pF0005BSdel2-6 (2-6) were prepared and assayed by S1 nuclease analysis. Four of 12 monoclones were positive for pF0005BS. Five of 12 and 1 of 12 were positive for 2-10 and 2-6 respectively. The S1 analysis of these cell lines (Figure 3-25) was repeated 2 times and the average expression from both pF0005BS and 2-10 was identical (1.5 ± 1.4). This value represents only the absolute amount of the human S1 protected fragment, as measured by the densitometer, averaged for each construct. Since only a single monoclone was positive for 2-6 it was not included in the analysis. The mouse H4 S1 protected fragment presented an unusual pattern even upon repetition and was not included in the analysis. The control S1 nuclease analysis of C127 RNA worked well (Figure 3-25) but the sample C127 protected bands were always

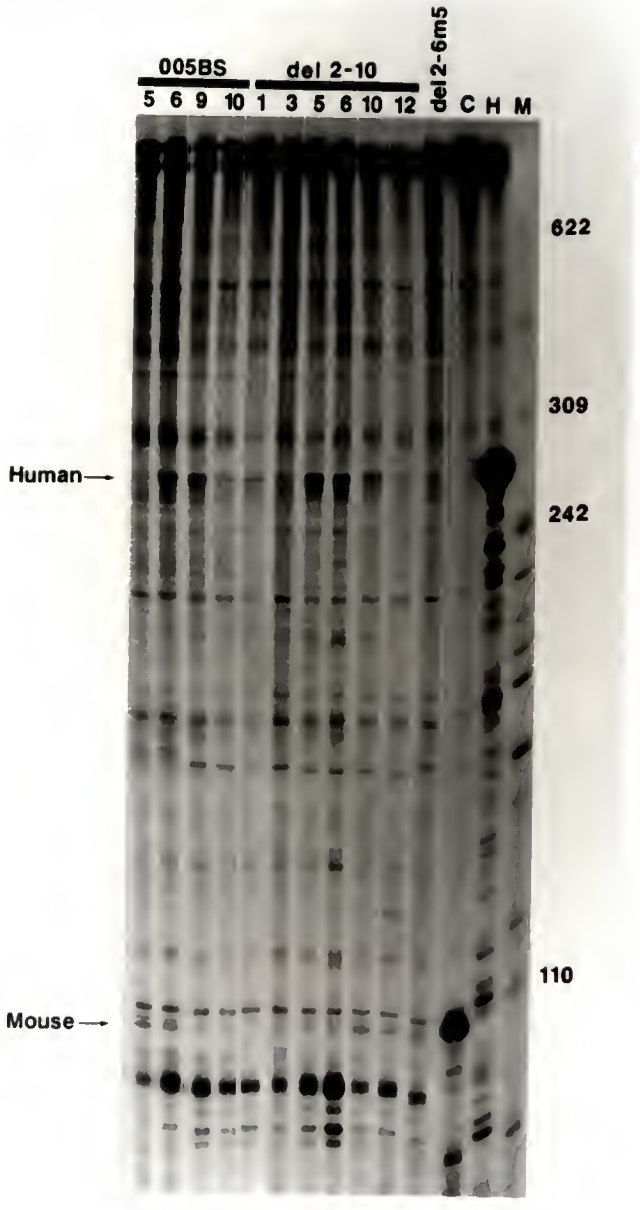


Figure 3-25 S1 analysis of pFO005BS and Exonuclease III deletions.

S1 nuclease assays were done as described in Materials and Methods. The clone number of each cell line is denoted above the lane. M, pBR322 digested with HpaII and labelled with α-^{32}P-dCTP and Klenow fragment. H, HeLa total cellular RNA. C, C127 total cellular RNA. The human (280 nt) and mouse (110 nt) S1 nuclease protected fragments are denoted at the left.

lower or not detectable even when the human H4 protected fragment was easily seen. We have noticed this occasionally, but never in so many samples at once. Occasionally, high copy number cell lines or the HeLa total cellular RNA control have exhibited a similar pattern of hybridization, but there has been no consistency.

The results of transient assays of pFO005BS, 2-10, and 2-6 were inconsistent due to quantitation problems with the plasmid DNAs that were not discovered until after the completion of the analysis. Even with these problems the results of the transient assays supported the idea that pFO005 and pFO005BS were the same with respect to the level of expression (data not shown). Even though more DNA was added in the transfection than originally thought, we were able to conclude that 2-10 (-335 bp) was not significantly different than pFO005BS or pFO005. Unfortunately, we were not able to make any conclusions about the 2-6 construct from the transient assays. We can only say that in stable cell lines it was expressed at a low level in the single monoclone of 12 that was positive. The results support the contention that removal of 80 bp (2-10) from pFO005 has little effect on the level of expression. The transcriptional analysis of these deletions has been repeated by Ken Wright with HeLa nuclear extracts <u>in vitro</u>. He has reached similar results to those presented here (personal communication).

Enhancer Element

Dr. Sherron Helms of our laboratory had previously identified the distal EcoRI/EcoRI fragment (-6.0 to -7.5 kb), designated pFO116, of the λ HHG41 clone as a possible enhancer element (Helms et al., 1987).

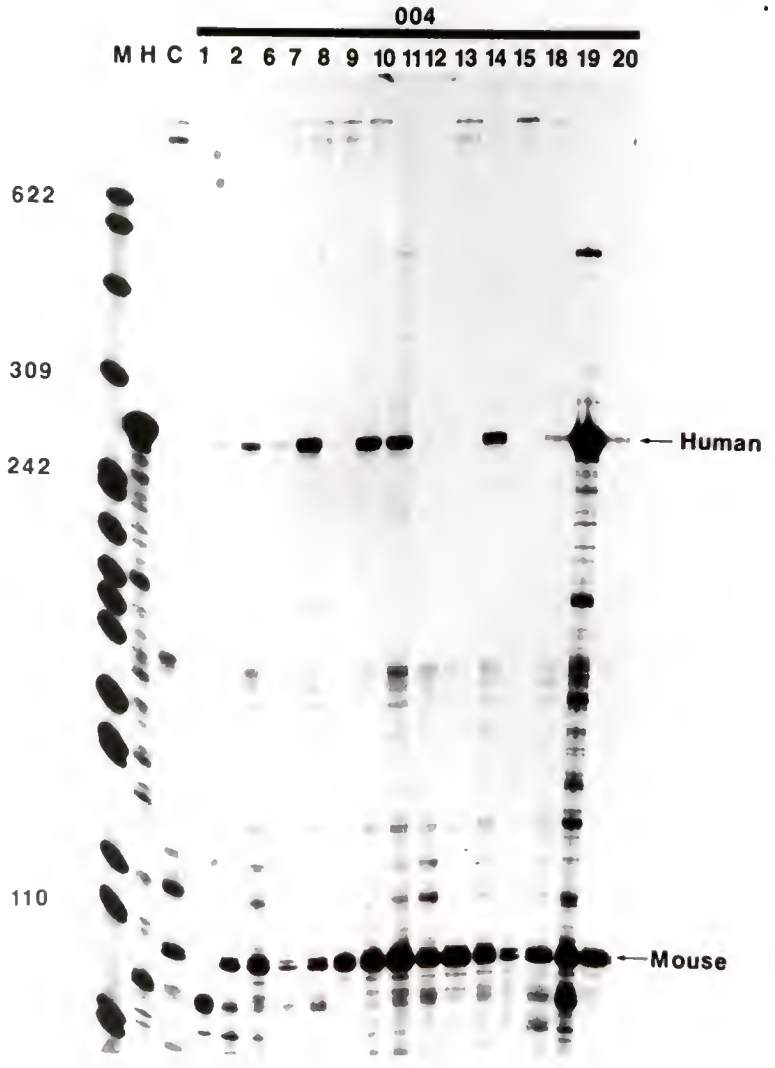


Figure 3-26 S1 nuclease analysis of pFO004 monoclonal cell lines.

S1 nuclease assays were performed as described in Materials and Methods. The number of each clone is displayed above the lane and below the construct designation line. M, pBR322 digested with HpaII and labelled with α-^{32}P-dCTP and Klenow fragment. H, HeLa total cellular RNA. C, C127 total cellular RNA. Both human (280 nt) and mouse (110 nt) H4 protected fragments are designated. Pertinent markers are noted.

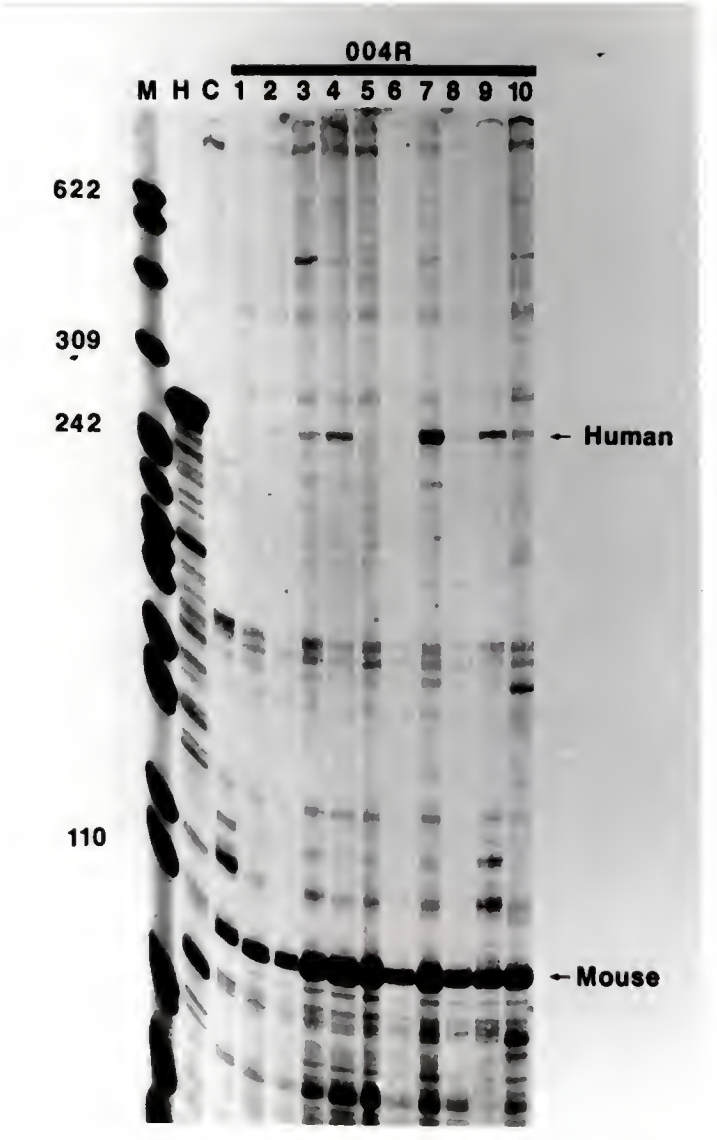
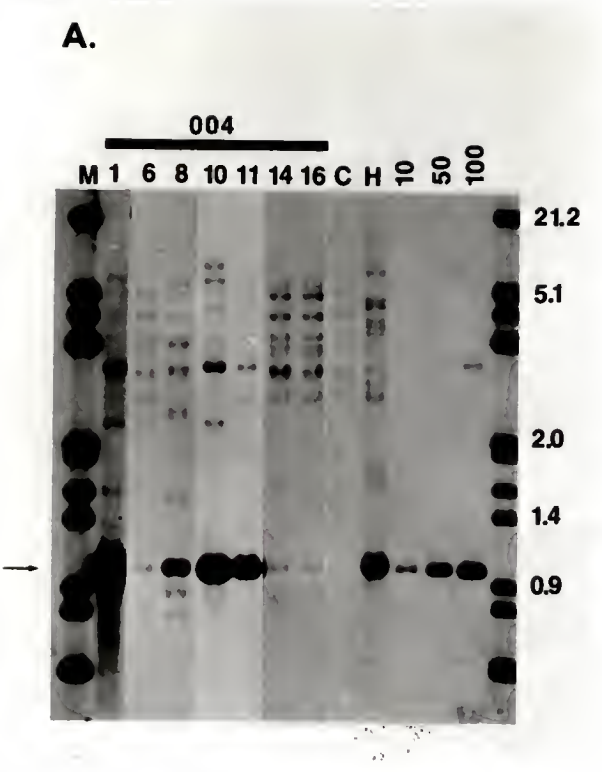


Figure 3-27 S1 nuclease analysis of pF0004R monoclonal cell lines.

S1 nuclease assays were performed as described in Materials and Methods. M, pBR322 DNA digested with HpaII and labelled α-^{32}P-dCTP and Klenow fragment. H, HeLa total cellular RNA. C, C127 total cellular RNA. Clone numbers (1-10) are designated above each lane. Both human and mouse H4 protected fragments are noted at the right.

Figure 3-28 Copy number analysis of pFO004 monoclonal cell lines.

Southern blot analysis was performed as described in Materials and Methods. 10 μg of DNA from each cell line were analyzed with nick translated EcoRI/XbaI fragment from pFO002. A. pFO004 cell line DNA probed with H4 sequences. B. The histone probe was removed and the blot was reprobed with the mouse 18S ribosomal fragment. Densitometry of the 1070 bp band specified by the arrow in A and the 18S ribosomal band in B permitted quantitation of the copy number through normalization to the amount of DNA actually loaded and transferred as described in the Materials and Methods. The figure in A is a composite of several exposures that reflects the actual copy number and accounts for original quantitation errors. The plasmid controls for quantitation are labelled 10, 50 and 100 designating the number of pg loaded. C, C127 cellular DNA. H, HeLa cellular DNA. M, λ DNA digested with EcoRI and Hind III and labelled α-^{32}P-dCTP and Klenow fragment. Clones are designated with their number above each lane. Nonessential lanes in B have been omitted.


A.
004
M 1 6 8 10 11 14 16 C H 10 50 100
21.2
5.1
2.0
1.4
0.9
B.
004
M 1 6 8 10 11 14 16 C H
18S
1.3

She demonstrated that linkage of this fragment to the 3' end of the CAT gene in pSV1CAT (Gorman et al., 1982) increased the expression of the CAT gene 4 to 5 times in transiently transfected HeLa cells. Enhancers can be located at considerable distances from the gene that they effect. The chicken lysozyme gene enhancer is located 7 kb upstream of the gene (Theisen et al., 1986). To investigate this result further a series of constructs were made and assayed in stable cell lines. The constructs pFO004, pFO04R, and pFO006 are depicted schematically in Figure 3-1 and were constructed by standard cloning procedures. pFO004 fused the pFO116 fragment to the pFO108P construct in the genomic orientation. The construct pFO004R was made to reverse the orientation, and pFO006 linked the 500 bp EcoRI/XbaI fragment of pFO116 to pFO108X.

The constructs were transfected into mouse C127 cells and selected for the growth of monoclonal cell lines. The S1 nuclease analysis of pFO004, pFO004R, pFO006 and the control cell line pFO108X are presented in Figures 3-26, 3-27, and 3-15 respectively. The expression data are presented in Table 3-2. The only cell lines with significantly higher levels of expression than pFO108X contained the pFO004 construct. To determine if this was truly the result of an enhancement or a phenomenon of copy number, the latter was determined by Southern blot analysis (Figure 3-28) as described previously in Materials and Methods. When the pFO004 cell line copy numbers were included in the expression/copy ratio, the level of expression dropped to control (pFO108X) level. Since neither pFO004R nor pFO006 had a significant difference in expression from pFO108X, the copy number for these cell

lines was not determined except for pFO006m3. This cell line presented the unusual mouse S1 nuclease protected fragment seen with the pFO005BS cell lines and therefore had high expression. The copy number was determined to exclude the possibility that this was an enhancer effect. pFO006m3 was included on the pFO108X copy number blot (Figure 3-16), and from this blot it was determined that pFO006m3 has approximately 30 integrated copies, which accounted for the higher level of expression.

The lack of an enhancer effect by pFO116 was surprising in light of the previous demonstration of the effects on CAT gene expression. The sequence for the entire pFO116 fragment was determined by Ken Wright and Urs Pauli of our laboratory and the proximal EcoRI/XbaI fragment contains three sequences with similarity to the consensus core enhancer element (5'-TGTGGAAA-3') as described for the Ig heavy chain and SV40 enhancers (Wasylyk and Wasylyk, 1987; Khoury and Gruss, 1983). The presence of this sequence has been shown not to be solely sufficient or necessary for enhancer activity in the IgH enhancer (Wasylyk and Wasylyk, 1987; Kadesh et al., 1986). The reasons for a lack of activity in mouse C127 cells, and activity in HeLa cells, is purely speculative. Certainly differences in the proteins that interact with enhancers in different tissue types have been documented (Maniatis et al., 1987; Davidson et al., 1986). The evidence from the stable cell lines we have prepared does not support the idea that the pFO116 fragment enhances or augments the expression of the FO108 H4 histone gene when stably integrated in a mouse cell. The fact that the pFO004 monoclonal cell lines had such a high average copy number has been

investigated further in Chapter 4 with respect to specificity and mode of integration. Ken Wright of our laboratory has also demonstrated that the pFO116 fragment is unable to enhance the transcription *in vitro* of the human H4 gene (personal communication).

The contribution of promoter sequences to the expression of the FO108 H4 histone gene has been determined in stable cell lines. Initially this approach appeared to be the most accurate way to determine functionality; however, in retrospect there are a number of variables which can not be accounted for. A brief comparison of the results from the transient assays done to assess the negative regulatory element hypothesis indicates the heterogeneity that can occur in the results and their interpretation as a result of the methodology utilized to perform the experiment. We originally thought that the mouse H4 S1 nuclease probe would be the ideal internal control, but subsequently we have realized that it has faults for which we cannot correct in our interpretation of the results. The possibility exists that limiting transcription factors are present in only sufficient amounts to transcribe the mouse H4 genes present in the cell. The introduction of the human H4 genes into the genome of the mouse cell likely disturbs this equilibrium. The possibilities for misinterpretation are considerable. If, at low copy number, the human H4 genes do not effectively compete for mouse transcription factors then we have probably underestimated their relative expression. At high copy number it is quite apparent that the expression/copy ratio decreases. Based on what we have presented here and later in Chapter 4 we will formally assess the results of the transcription data in

Chapter 5 with respect to the low copy number cell lines only. Although this limited my data base it appeared to be the only reasonable way to proceed in order to fairly evaluate the data we have collected.

Nuclear run-on analysis of H4 transcription

This section of the results is added purely as a note to those who might try similar experiments as described below. None of the experiments we have described above directly assess the level of transcription. Differences in the 5′ region of the promoter were assayed in log phase cells under the assumption that the mRNA was the same for all constructs, therefore any differences in the level of mRNA were a reflection of transcription. This interpretation is fine and holds up reasonably well when deletion constructs are compared to one another. However, to examine transcription directly it is necessary to eliminate the mRNA stability variable in histone gene metabolism. Our laboratory has utilized nuclear run-on transcription to identify the time and extent of human histone gene transcription during the cell cycle (Baumbach et al., 1987). We felt that our monoclonal cell lines would be ideal candidates for such an analysis and that we could determine the region of the promoter responsible for the 3-5 fold increase in the level of transcription during S-phase of the cell cycle. Briefly, nuclei were isolated from the cells at 4°C and transcription allowed to continue in the presence of α-^{32}P-UTP for 30 minutes. The labelled RNA was purified and used to probe blots that had plasmid DNAs immobilized (excess DNA hybridization). In short, regardless of the temperature, salt concentration, or aqueous state of

the hybridization reaction we were able to observe only mouse H4 mRNA cross hybridization to the human H4 plasmid DNA (data not shown). An alternative approach to the detection problem was tried -- what we called a "reverse S1 analysis". The labelled nuclear run-on transcripts were incubated with cold probe RNA made from the T3 promoter of a Bluescript clone of pFO002. This was then digested with S1 nuclease, electrophoresed, and visualized as usual. As a control, the probe pFO002 RNA was labelled with α-^{32}P-UTP and hybridized to HeLa total cellular RNA. The control worked well, but the test reaction was only a smear (data not shown). This result had previously been predicted by my outside examiner Dr.Barbara Sollner-Webb. She felt that the technique would not work because of stable double stranded ribosomal RNA that would be labelled and obscure the histone signal. We decided that unfortunately this approach was not possible in our system.

CHAPTER 4

PLASMID INTEGRATION SITES, INTEGRITY AND PROTEIN/DNA INTERACTIONS

One goal of modern molecular biology is to understand the molecular events that occur during the integration of exogenous DNA into the chromosome of a cell. These processes have been examined in detail by several investigators and are important for the study of biological problems in eukaryotic cells (Loyter et al., 1982, Perucho et al., 1980, Folger et al., 1982, Lin and Sternberg, 1984). The problem of what happens to the DNA molecules once they enter the cell is intriguing as it gives a glimpse of the complicated recombinational processes that occur inside the cell. Loyter et al. (1982), demonstrated that there was a limit to the amount of DNA a plate of cells could take up, and that only a small percentage of the DNA that entered the cytoplasm subsequently entered the nucleus. Previous work by Perucho et al. (1980, 1981) demonstrated that as foreign DNA (e.g. plasmid DNA with a gene of interest) entered the nucleus of a cell it became recombinationally active. Because there is usually little or no homology between the foreign DNA and the cellular DNA the first recombination events that occur are between the plasmid DNA molecules and carrier DNA. Cointegrates form in the nucleus shortly after the introduction of the DNA into the cells. These very large circular molecules contain many plasmid molecules arranged in a head-to-tail

manner. When integration into the host chromosome does occur, a large number of plasmid molecules are likely to integrate stably in a head-to-tail fashion at a single location (Perucho et al., 1980). For my purposes it was necessary to assess the integrity of the integrated histone deletion constructs, the structural relationship to the cotransfected plasmid pSV2neo, and the mode of integration.

Integrity of Flanking Sequences

To assess the intactness of the proximal flanking sequences, we examined the copy number blots of the constructs with 210 bp or less of 5' flanking sequence. In the constructs J67 (-47 bp), J56 (-73 bp), J50 (-100 bp), K8 (-155 bp), L14 (-185 bp) and pFO108A (-215 bp) (Figures 3-4, 3-6, 3-10, 3-14) the EcoRI/XbaI fragment represents the entire coding and 5' flanking sequences. The EcoRI/XbaI digest was originally chosen because it is a fragment common to all the constructs used in the study. It was easily determined by inspection of these Southern blots that all or nearly all of the H4 constructs were integrated in a manner that permitted detection of the human H4 insert sequences of the original plasmid. The integrity of longer constructs such as pFO003 and pFO004 was not measurable this way. To assess the integrity of the pFO003 flanking sequences the genomic DNA from a polyclonal cell line pFO003p3 was digested with XbaI. This digestion defines the entire 7.5 kb insert. The restricted DNA was electrophoresed on a 1% agarose gel, blotted, and probed with the EcoRI/XbaI fragment from pFO002. The results, presented in Figure 4-1a, lane 4, demonstrate the predominance of a 7.5 kb band that corresponds to the presence of the entire pFO003 insert. Genomic DNA

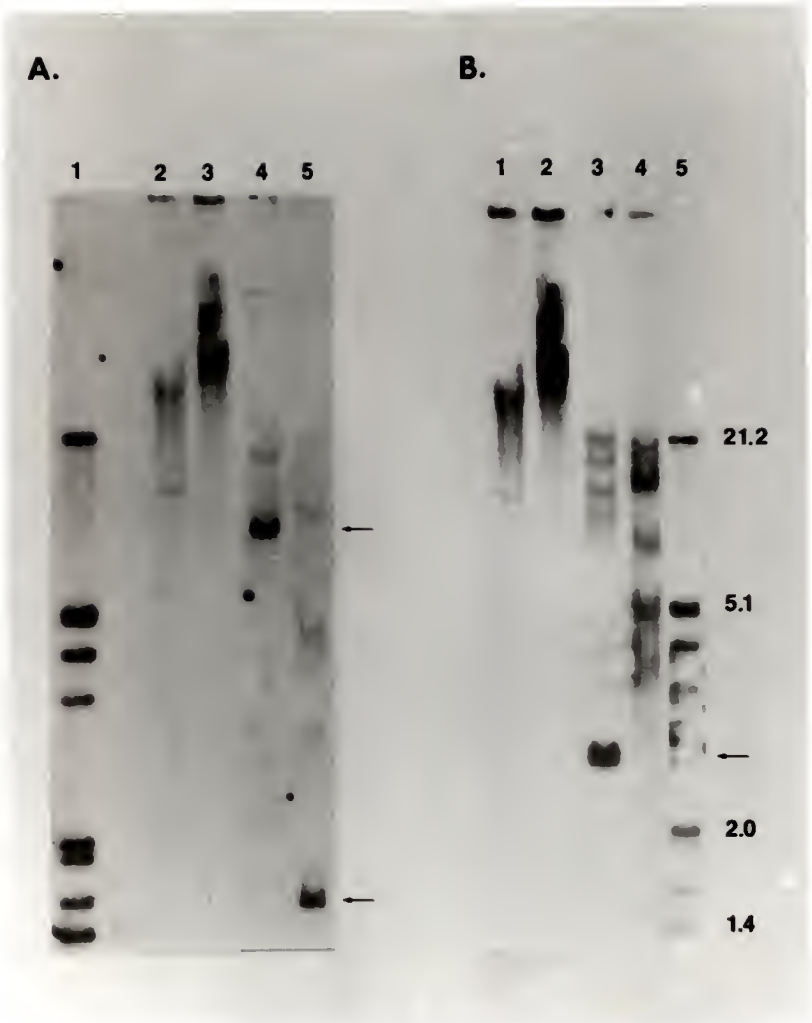


Figure 4-1 Southern Blot analysis of genomic DNA from polyclonal cell lines: assessment of flanking and coding sequence integrity in pFO003 and pFO004.

Ten micrograms of genomic DNA from polyclonal cell lines pFO003p3 and pFO004p2 were digested with either XhoI or XbaI, electrophoresed, blotted, and probed as described in Materials and Methods. A. Lane 1, λ DNA digested with HindIII and labelled with α-^{32}P-dCTP and Klenow fragment; Lane 2, XhoI digested pFO003p3 DNA; Lane 3, XhoI digested pFO004p2 DNA; Lane 4, XbaI digested pFO003p3 DNA; Lane 5, XbaI digested pFO004p2 DNA. The blot was probed with the histone H4 EcoRI/XbaI fragment purified from pFO002 and nick translated as described in Materials and Methods. In lanes 4 and 5 the expected size fragments are noted with arrows at the right. B. Lanes 1-4 are the same as 2-5 in A. Lane 5, λ DNA digested with HindIII and labelled with α-^{32}P-dCTP and Klenow fragment. The blot in B was probed with nick translated pUC8 DNA. The 2.7 kb band in lane 3 (pFO003p3) is linear pUC8.

from a polyclonal cell line of pFO004 was digested with XbaI and it was established that a large percentage of the 1.6 kb XbaI/XbaI fragment that includes the coding region and much of the flanking sequences was detectable as an intact fragment (Figure 4-1a, lane 5). To determine that the DNA was indeed integrated we digested both pFO003p3 and pFO004p2 with XhoI, an enzyme that has no sites within either plasmid. Figure 4-1, lanes 2 and 3, demonstrate that when the DNA is digested with XhoI almost all of the hybridization to the human histone probe is in the region of the blot that corresponds to very high molecular weight DNA. Evidence for tandem integration was found when the blot in 4-1a was reprobed with pUC8 DNA. In Figure 4-1b, lanes 1 and 2 still demonstrate high molecular weight DNA as expected. Lane 3 has a predominant 2.7 kb band that is probably pUC13. The fact that both the pFO003 insert (7.5 kb) and vector (2.7 kb) bands were so readily detectable was indicative of tandem integration. Unexpectedly the pFO004p2 DNA did not have a similar 2.7 kb band. Instead there was a heterogeneous pattern of hybridization to the pUC8 DNA observed in Figure 4-1b, lane 5.

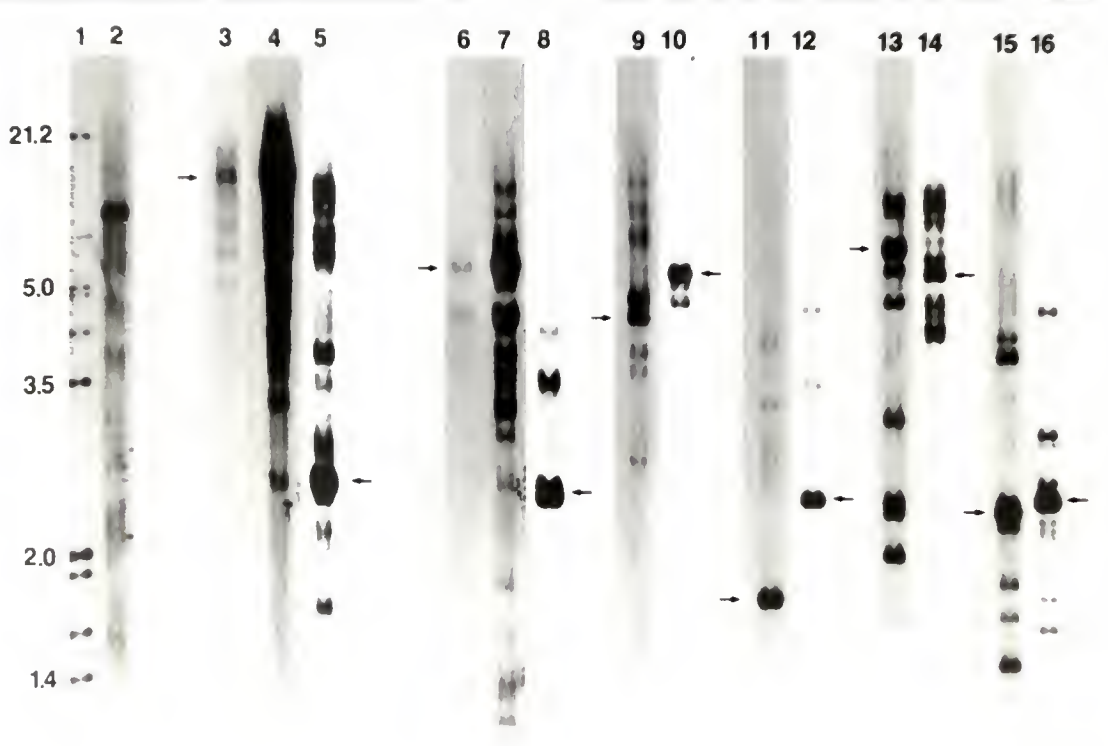
These experiments were pursued further to establish the mode of integration that had occurred in the monoclonal cell lines. This information would allow one perhaps to understand how, or if, the arrangement of histone insert sequences with respect to each other affects expression. Our concern has been how to interpret the copy number data with respect to expression. The possibilities are considerable that tandem integration, for example, might "protect" internal integrates from chromosomal effects in <u>cis</u>. Is it possible to

assume that all the genes integrated in a cluster are going to function equally well? When polyclonal cell line DNA from 108Ap2 was digested with PstI and analyzed by Southern blotting, it was evident that a single band of 6.2 kb was present and corresponded to full length plasmid DNA (data not shown). Therefore integration appeared to have occurred in a tandem fashion. The detection of such a high percentage of the EcoRI/XbaI fragment in the copy number blots mentioned above also demonstrated that tandem integration was probably the pathway utilized.

We analyzed several other monoclonal cell lines with different restriction enzymes and Southern blot analysis to establish that tandem integration was a general phenomenon. In Figure 4-2, lanes 3 and 4 demonstrate that when genomic DNA from the monoclonal cell line pFO003m1 was digested with PstI a predominant 10.2 kb band was detected following hybridization with an oligo-labelled 3' noncoding XbaI/HincII fragment of pFO002. Lane 3 is just a lighter exposure of lane 4. We concluded that tandem integration was apparently the mechanism used by most constructs. This human H4 histone gene 3' probe permitted detection of only human histone sequences since it contained no coding region. In Figure 4-2, lane 9, pFO005m5 genomic DNA was digested with BamHI, again an enzyme that linearizes the construct. Two bands were detected with the histone 3' probe, linear pFO005 (4.3 kb) and a slightly higher band that was not identified. This pointed toward tandem integration, and limited heterogeneity of integration sites. Digestion of pFO108Am10 genomic DNA with BamHI (Figure 4-2, lane 13) demonstrated more heterogeneity although the correct size

Figure 4-2 Southern Blot analysis of monoclonal cell line integration pattern and location of pSV2neo sequences.

This figure is a composite of the same blot that has been probed with two different DNA fragments. The blot was first probed with a 3' fragment from the FO108 H4 histone gene as described in the text, and Material and Methods. This probe was removed and the blot was probed for a second time with a fragment that contains the SV40 enhancer. The complementary lanes from each analysis have been placed next to each other to facilitate comparison of the data. Lanes 2-4, 6, 7, 9, 11, 13, and 15 were all probed with the 3' histone H4 fragment. Lanes 5, 8, 10, 12, 14 and 16 were probed with the SV40 fragment. Lanes: 1, λ DNA digested with HindIII/EcoRI and Klenow labelled; 2, HeLa DNA digested with XbaI and a 7.5 kb band is detected. Lanes 3 and 4, pFO003m1 DNA digested with PstI. The 10.2 kb linear pFO003 molecule is denoted at the left (3 is a shorter exposure of lane 4). Lane 5, pFO003m1 DNA digested with PstI and the 2.3 kb PstI pSV2neo fragment is indicated at the right by an arrow. Lanes 6 and 7, pFO004M11 DNA digested with Pst 1. A 5.7 kb linear band is detected and indicated (lane 6 is a shorter exposure of lane 7). Lane 8, pFO004m11 DNA digested with PstI. The 2.3 kb PstI fragment of pSV2neo is denoted. Lanes 9 and 10, pFO005m5 DNA digested with BamHl. In lane 9 a 4.3 kb linear pFO005 band is denoted. In lane 10, the linear 5.5 kb pSV2neo band is indicated. Lanes 11 and 12, pFO005m5 DNA digested with PstI. In lane 11 a 1.7 kb band corresponding to the entire pFO005 insert is detected. In lane 12, the 2.3 kb PstI fragment of pSV2neo is noted. Lanes 13 and 14, pFO108Am10 DNA digested with BamHI. In lane 13 the 6.2 kb band of linear pFO108A is noted. In lane 14 the 5.5 kb pSV2neo band is noted. Lanes 15 and 16, pFO108Am10 digested with PstI. In lane 15 the 2.2 kb band corresponding to most of the pFO108A insert is detected. In lane 16 a 2.3 kb pSV2neo band is detected as expected.


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
21.2
5.0
3.5
2.0
1.4

linear fragment was detectable (6.2 kb). When pFO005m11 and pFO108Am10 DNA were digested with PstI, a fragment of the expected size was readily detectable (Figure 4-2, lanes 11 and 15, respectively). These results were consistent with tandem integration, perhaps in several locations.

We noticed early in these studies that cell lines that contained the construct pFO004 had a heterogeneous pattern of integration. When pFO004m11 genomic DNA was digested with PstI, an enzyme that linearizes the construct, there were fewer linear molecules (5.7 kb) detectable, Figure 4-2, lanes 6 and 7. Lane 6 is a lighter exposure of lane 7. This increase in the heterogeneity of integration was associated with the presence of an Alu repeat sequence in the 5' flanking region of this construct. We examined the copy number data and calculated the average copy number of each type of cell line (Table 3-2) and were able to correlate the presence of repeated sequences with increased average copy number. Previous work on repeated sequences associated with histone gene clusters (Collart et al., 1985) had demonstrated the presence of a strong Alu repeat in the most distal EcoRI/XbaI fragment (-5.5 to -6.5 kb) of the putative H4 promoter sequences, and, to a lesser extent, minor repeated sequences located between the BamHI site (-1.65 kb) and the EcoRI site at -5.5 kb. The fact that the pFO003m1 cell line had a significant proportion of its DNA tandemly integrated as shown in Figure 4-2, lane 3, suggested that while the minor repeats located in its flanking sequence have contributed to increased copy number, they have not caused as much heterogeneity in the integration sites as the Alu repeat in pFO004.

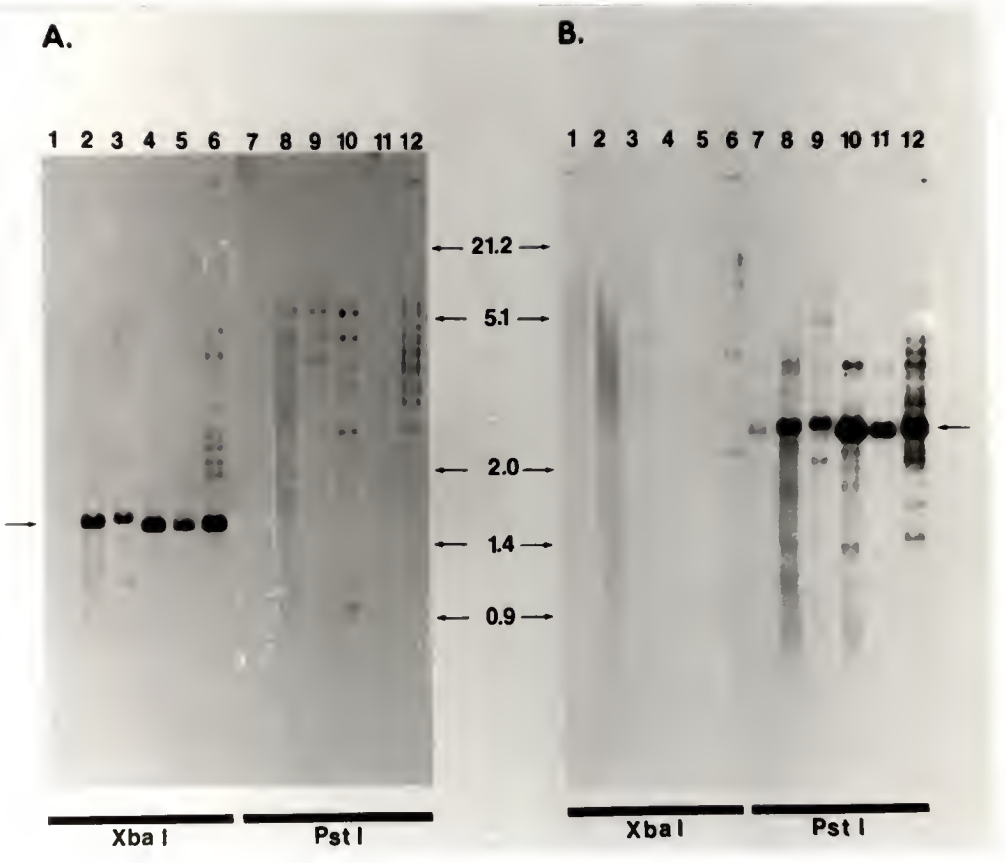


Figure 4-3 Specificity of pF0004 integration.

Southern blot analysis was done according to the procedures described in Materials and Methods. A. Six pF0004 cell lines: pF0004p1, pF0004m1. pF0004m8, pF0004m10, PF0004m11, and pF0004m19 were digested with either XbaI (lanes 1-6) or PstI (lanes 7-12). The blot was probed with oligo labelled XbaI/HincII fragment of pF0002 as described in Materials and Methods. The 1.6 kb insert band of pF0004 is designated to the left of the figure. B. The histone 3' probe was removed and the blot was reprobed with oligolabelled 264 bp SV40 enhancer fragment as described earlier. The lanes in B are identical to those in A. The position of marker fragments is designated between A and B in kilobases. The restriction enzyme digest is indicated below each lane.

The previous demonstration with the pFO004pl cell line that a significant portion of the insert was detectable and intact (Figure 4-1, lane 5) suggested that these contrasting results might be the consequence of specific integration. To determine whether the ability to detect the pFO004 insert fragment in genomic DNA was limited to the single polyclonal cell line we had examined, we repeated the experiment with the pFO004pl cell line and 5 monoclonal cell lines. Cell lines with reasonably high copy number were utilized to aid detection and assess the effect of integration on intactness of the flanking sequences. The results, presented in Figure 4-3a, lanes 1-6, demonstrate that in every cell line the 1.6 kb XbaI/XbaI fragment was detectable and constituted a considerable portion of the signal present in each lane. If the same pFO004 monoclonal cell line DNAs were digested with PstI and probed for the presence of linear pFO004 molecules (Figure 4-3a, lanes 7-12) there was heterogeneity in integration and very little linear (5.7 kb) pFO004 was detectable. If so much of the insert XbaI/XbaI fragment was detectable, and so little linear, there must have been an unusual integration event that occurred to give both results. It appeared a strong possibility that the Alu sequence in the 5′ flanking region was a site where specific integration might occur. To test this hypothesis we digested three of the pFO004 monoclonal cell line DNAs with NcoI. This enzyme has two sites of digestion, one at +280 bp and one in the very distal 100 bps of the 5′ flanking sequence (-1.6 kb, Figure 4-4b). This digest produces two DNA fragments of 1.9 kb and 3.8 kb. The Alu sequence is located in the 1.9 kb NcoI fragment. It is important to recall that

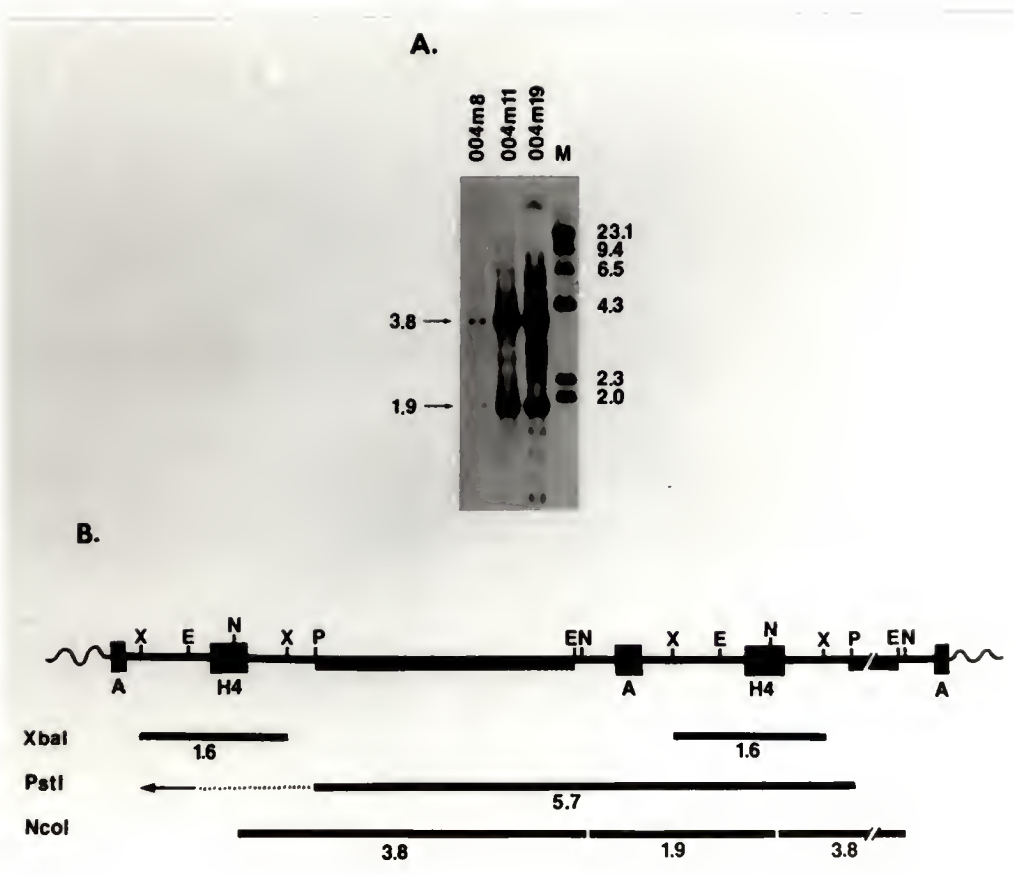


Figure 4-4 Southern blot analysis of pFO004 integration: NcoI digestion of genomic DNA.

The analysis was done as described in Materials and Methods. A. 10 μg of DNA from monoclonal lines pFO004m8, 11, and 19 were digested to completion with NcoI. The blot was probed with the EcoRI/XbaI fragment from pFO002. This probe detects both the 1.9 kb and 3.8 kb bands. B. Synopsis of the hybridization to the EcoRI/XbaI probe. This figure presumes that two pFO004 molecules have integrated tandemly head to tail through one of the Alu repeats. pFO004, when digested with XbaI, produces a homogeneous 1.6 kb band and 2 copies are detectable. Digestion with PstI produces a single detectable linear molecule of 5.7 kb and in this case one end fragment designated by the dotted line and the arrow. The data from part A of this figure supports the fact that 2, 3.8 kb, and 1, 1.9 kb fragment would be detectable from this double integrate.

when the pFO004 monoclonal DNAs were digested with XbaI, uniformly a 1.6 kb fragment was detectable that indicated that the flanking sequences up to -715 bp were intact and integration had not occurred between the two XbaI sites. The only other regions available for integration were the XbaI/EcoRI fragment (-750 to -1750 bp) and the pUC13 vector sequences. Because all constructs share similar vector sequences it was unlikely to be this region that differentiated the pFO004 construct from others in integrative mode. The pFO004 NcoI digestion experiment was probed with the EcoRI/XbaI fragment of pFO002 which detects both the 1.9 and 3.8 kb NcoI fragments. The results of the NcoI digestion Southern blot are presented in Figure 4-4a. Several exposures were scanned densitometrically to determine the ratio of 3.8 kb to 1.9 kb fragment. In the three cells lines the ratio of the 3.8 kb and 1.9 kb NcoI fragments was approximately 2:1.

To explain this and previous results our current hypothesis is that the pFO004 plasmid DNA integrated through the Alu sequence in no more that two or three copies per integration site. This hypothesis explains, as diagrammed in Figure 4-4b, that when two copies of pFO004 are integrated through the Alu sequence: 1) the XbaI/XbaI fragments (there are 2) are both detectable, 2) only one of the two integrated constructs is detectable when the PstI digestion is done, and 3) the NcoI digestion produces two 3.8 kb and one 1.9 kb fragments as seen experimently (Figure 4-4a). Althoughnot conclusive it suggests some preferential integration via the Alu sequence. This specificity of integration through the Alu repeat accounts for the heterogeneity in integration sites observed. Previously it was thought that integration

of plasmid DNA, in a single eukaryotic cell, occurred first at the cointegrate stage, then at a single site in the chromosome (Perucho et al., 1980). This is plausible because usually there is relatively little homology between the transfected DNA molecules and the cellular DNA. However, it is apparent from our results that repeated sequences, such as the Alu repeat, which are well conserved from species to species may mediate specific and higher levels of integration than normally possible.

Location of pSV2neo Plasmid Sequences

The second point to be addressed in these experiments was whether the pSV2neo plasmid was located in the proximity of the human histone H4 gene constructs. In order to create the cell lines we have used in this study, it was necessary to cotransfect with the plasmid pSV2neo. Our primary concern was to establish to what extent the SV40 enhancer might associate with the histone promoter deletion constructs and affect the expression of the H4 constructs in a *cis* manner. Since there is similarity between the pBR322 portions of these various plasmids we investigated the possibility that pSV2neo and histone deletion plasmids were located adjacent to each other.

An early observation with regard to this problem was that constructs such as J67 and other short deletion constructs of the H4 promoter demonstrated little or no transcription when integrated stably. This result, probably more than any other, demonstrated that the pSV2neo plasmid had little or no influence on expression of the cotransfected histone plasmids. It was reasonable to suppose that the integration of the human H4 histone genes occurred at a sufficient

distance from the influence of any endogenous strong promoter effects. It was also unlikely that the pSV2neo cotransfected molecules had any substantial effect on the expression of the transfected H4 histone genes.

To initially address the location with respect to human histone H4 sequences of the pSV2neo plasmid in the monoclonal cell lines, we reprobed monoclonal cell line Southern blots shown in Figure 4-2 with an oligo-labelled EcoRI/EcoRI fragment that contained the entire SV40 enhancer sequence. A pUC8 clone of the 264 bp EcoRI/EcoRI fragment was kindly provided to me by Gerard Zambetti of our laboratory and contains both 72 bp repeats (originally derived from pDG014, a gift of Dr. Sherman Weissman). Figure 4-2 is a composite of identical lanes probed with either the histone 3' probe as detailed earlier or the SV40 enhancer fragment. We felt it would be easier for comparison if the lanes were placed adjacent to each other instead of on separate figures. In Figure 4-2 lanes 5 (pFO003m1), 8 (pFO004m11), 12 (pFO005m5), and 16 (pFO108A) are identical to the adjacent lanes 4, 7, 11, and 15. A PstI digest of pSV2neo produces three fragments and the SV40 probe detects the 2.3 kb fragment that contains part of the neomycin resistance gene, the SV40 promoter/enhancer, and some pBR322 sequence. When genomic DNA is digested with BamHI the pSV2neo DNA is linearized, and if tandemly integrated, a 5.5 kb band should be detectable. In Figure 4-2, lane 5, the pFO003m1 DNA cut with PstI demonstrated a prominent 2.3 kb band as expected, but also has bands in the region of the histone signal detected previously (10 kb) in lanes 3 and 4. The ability to detect a substantial amount of both the 10.2 kb

pFO003 DNA and the 2.3 kb pSV2neo fragment suggests that there is not a substantial mixing of the two molecules in the pFO003 integration site. The smaller fragments detected in lane 4 probably represent "end" fragments of each integration event. An end fragment is detected because it is the most distal plasmid sequence on either side of the integration event. In the case of pFO003ml, the DNA is cut with PstI. The pFO003 DNA molecules integrated at each end of the tandem array will be subject to cutting internally with PstI once, and at some unknown distance into the cellular DNA at the next available PstI site. Since this next PstI site is of an undetermined location on both ends of the integration event, for every integration site there will usually be two end fragments of unknown length detectable. The number of end fragments can indicate the number of different sites into which the construct has integrated. pFO003ml has 10 or more fragments in addition to the main band at 10 kb. This could be interpreted as reflecting 5 integration sites in this monoclonal cell line or perhaps the inclusion of pSV2neo between tandemly repeated pFO003 molecules causes periodic interruptions.

The digestion of pFO004mll with PstI (Figure 4-2, lane 8) also demonstrates that the 2.3 kb pSV2neo fragment is detectable and constitutes a considerable portion of the signal in lane 8. A comparison of this lane hybridized to histone 3′ sequences (lane 7) and hybridized to the SV40 enhancer fragment demonstrates that very few of the bands detectable with the histone sequence probe are also detected with the SV40 enhancer probe. This lack of congruity pointed to some separation of the pSV2neo and pFO series plasmids upon integration.

The construct, pFO005, when digested with BamHI in Figure 4-2, lane 9, yields a 4.3 kb linear fragment when hybridized to histone 3' flanking sequences. When hybridized to the SV40 fragment, a 5.5 kb band corresponding to linear pSV2neo is detected (lane 10). The ability to detect a majority of the pSV2neo DNA as a linear molecule confirms the idea that in many instances the pSV2neo plasmid has integrated primarily in a site apart from the histone constructs. The band above the 4.3 kb (lane 9) and below 5.5 kb (lane 10) apparently contains both histone and SV40 sequences. The construct pFO108Am10, when digested with either BamHl (linearizes construct, 6.2 kb) (lane 13) or PstI (2.2 kb fragment) (lane 15), and probed with the histone 3' sequences resulted in the detection of many fragments. When hybridized with the SV40 enhancer fragment, there is a 5.5 kb band in lane 14 and a 2.3 kb band in lane 16, along with several additional bands, both larger and smaller. In some instances there is identity between the fragments detected by the two probes, so it is possible they are located in close proximity or linked on restriction fragments.

Still, at this point it was difficult to determine the relationship between the two transfected plasmids, histone and pSV2neo. It was apparent that in some cases there was a reason to believe that the two plasmids were not completely mixed during the integration events. To determine the relationship in a different way, we reprobed the blot in Figure 4-3a with the SV40 enhancer fragment after removal of the histone probe. The idea in this experiment was that the pSV2neo plasmid contains no XbaI restriction sites. Therefore, the digestion with XbaI, which released greater than 90% of the pFO004 sequences as a

1.6 kb band, should determine whether there was any mixing between the pSV2neo and pF0004 plasmid upon integration. The presence of a 5.5 kb or larger band would be indicative of mixing and release upon XbaI digestion. As can be seen in Figure 4-3b the XbaI digested pF0004 monoclonal cell lines (lanes 1-6) hybridized to the SV40 enhancer fragment in a diffuse manner and primarily in the upper region of the blot that was indicative of large DNA molecules. A few bands were detectable in the pF0004m19 cell line and it should be noted that this cell line has a very high copy number and a great deal of heterogeneous integration. The same pF0004 monoclonal cell lines when digested with PstI (Figure 4-3b, lanes 7-12) demonstrated that the pSV2neo sequences are present and detectable as a 2.3 kb band. The pF0004 monoclonal cell lines digested with XbaI and probed with the SV40 sequences suggest that the pSV2neo plasmid DNA is not interspersed in the integrated pF0004 plasmid DNAs. If the pSV2neo plasmid had been released by XbaI digestion we would have expected a strong band(s) in the high molecular weight region of the blot. The diffuse hybridization throughout the lane is somewhat confusing and unfortunately a C127 DNA control was not included on this gel. There is the possibility that the DNA fragments that contained the pSV2neo plasmid molecules were very large and did not transfer well from the gel.

Given the facts presented and known about enhancers, particularly the SV40 enhancer, it seems reasonable to conclude that this potent enhancer has little or no effect on the human H4 histone sequence integrated in these mouse cells. Because of the intensity of the

pSV2neo bands (2.3 kb) detectable in Figure 4-2 and Figure 4-3 it was likely that the copy number of the pSV2neo plasmid in these cell lines was very high. This was certainly the result of integration and amplification under the selective pressure of G418. Because of the selective pressure under which these cell lines were grown it was impossible to determine the absolute pSV2neo copy number originally present in the cell.

Compatibility of Mouse and Human Regulatory Proteins and Sequences

Examination of the copy number data presented in Table 3-2 revealed that as the copy number of a cell line increased the expression/copy decreased. This is graphically detailed in Figure 4-5 where several cell lines have been compared to one another for this effect. The obvious trend was typified by pFO005. When cell lines with fewer than 5 copies are plotted the expression/copy was high (0.3), but when copy number rose above 5 the expression/copy ratio decreased dramatically. Although the expression/copy ratio for the other constructs presented was generally lower than for pFO005, the decrease with increased copy number was still apparent. This effect presented several problems: 1) is it then appropriate to analyze only the low copy number cell lines for differences from construct to construct? and 2) does this indicate that the human and mouse H4 genes are in competition with each other for necessary transcription factors? In chapter 3 we alluded to the fact that there was a competition phenomenon. At that point we interpreted the pFO108A and pFO005 data in the context of copy number.

To determine whether these concerns were valid, we performed an analysis of the protein/DNA interactions in the 5′ promoter sequences

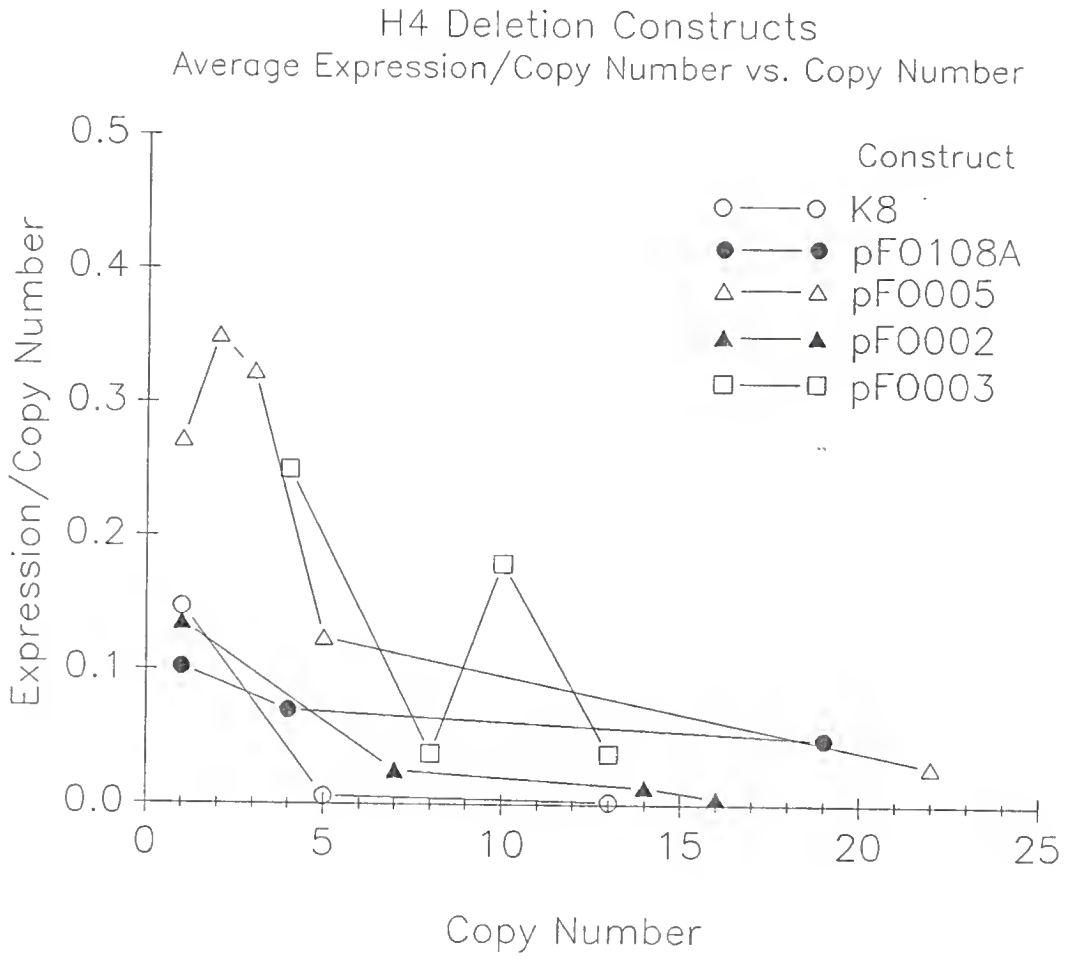


Figure 4-5 Effect of cell line copy number on the expression of the human H4 histone gene.

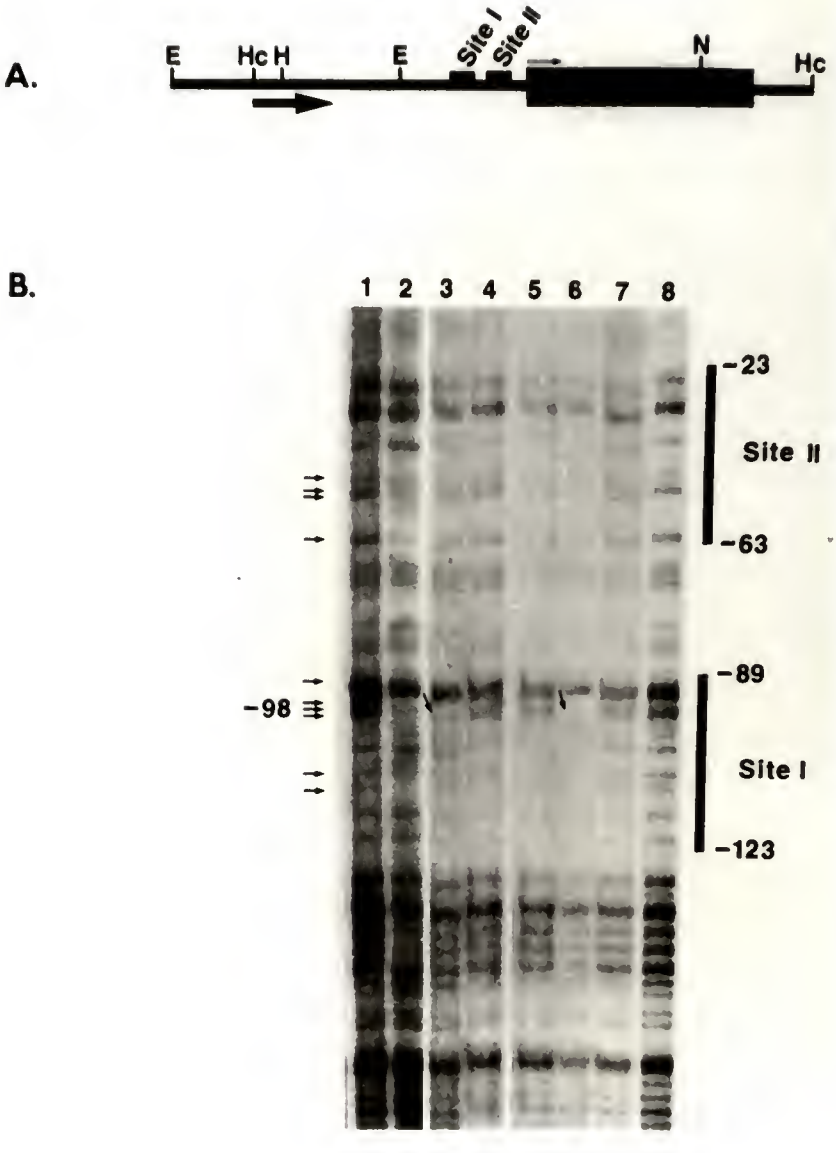
A plot of average expression/copy versus the cell line copy number. Data from Table 3-2 was averaged for K8, pF0108A, pF0005, pF0002, and pF0003. The average expression from all cell lines in a group with the same copy number are presented as single points. Most points are representative of the value for a single monoclone and not averaged with others. The legend in the figure designates each curve.

of the FO108 H4 histone gene. If promoter competition for transcription factors occurred then we felt it might be possible to detect the effect of high copy number on the binding of transcription factors.

In collaboration with Dr. Urs Pauli, of our laboratory, we characterized the protein/DNA interactions in the proximal promoter region of three monoclonal cell lines containing the construct pFO003. We were interested to know whether Site I and Site II were present in the proximal promoter of the human H4 histone gene when integrated in a mouse cell and if the protein/DNA contact points were the same. The pFO003 cell lines were chosen for several reasons. They had a wide range of copy number available and we felt that the extensive 5' flanking region (-6.5 kb) was more likely to assume a chromatin structure like that found in a human cell. Cell lines were grown until 80-90% confluent and treated with DMS in vivo as described in Materials and Methods. Genomic DNA from each cell line was prepared, digested with HincII, electrophoresed, and blotted as described in Materials and Methods. The filter with immobilized DNA was then hybridized with the 5' HincII upper strand probe (Figure 4-6a). This probe was used because the upper strand of the DNA contained 13 Gs strongly protected from DMS treatment whereas the lower strand contained only 3 minor protections (Pauli et al., 1987). All the G residues that exhibit protection are noted on the side of Figure 4-6. The boundaries of Site I and Site II are denoted to the right of Figure 4-6b. These were determined by Pauli et al. (1987) by DNaseI protection. Therefore, we were able to easily detect any differences

Figure 4-6 Genomic sequencing analysis: protein/DNA interactions in the proximal promoter of the FO108 H4 histone gene stably integrated into mouse C127 cells.

As described in Materials and Methods the genomic DNA from several different monoclonal cell lines of pFO003 was treated with DMS in vivo. The DNA was then purified, treated with piperidine, restricted with HincII, electrophoresed, blotted and probed with the upstream 5' HincII probe. A. Schematic diagram of the proximal region of the FO108 H4 histone gene. The single strand (HincII) probe that was utilized in these experiments is designated with the large arrow. Restriction enzyme sites are denoted as EcoRI, E; HincII, Hc; HindIII, H; NcoI, N. The large box is the H4 coding and leader sequence. Both Site I and Site II are designated above the diagram. B. Genomic sequencing analysis of protein contact points in Site I and Site II of the human H4 proximal promoter region. Lanes: 1, control HeLa DNA, purified, deproteinized, and then treated with DMS. 2, HeLa DNA that was treated in vivo to demonstrate the positions of Site I (-123 to -89 bp) and site II (-63 to -23 bp). At the left, the small arrows indicate the protein/DNA interactions as detected by DMS methylation interference (Pauli et al., 1987). Lane 3, pFO003m5 cell line DNA (copy number = 13) treated in vivo with DMS. The three G residues at approximately -98 to -100 bp are protected and denoted on the figure with an arrow. Lane 4, control deproteinized pFO003m5 DNA. Lane 5, HeLa control DNA, deproteinized and DMS treated. Lane 6, pFO003M6 (copy number = 20) DNA treated in vivo with DMS. The protected G residues at -98 to -100 bp are noted with an arrow on the figure. Lane 7, pFO003M1 (copy number = 140) DNA treated in vivo with DMS. The G residues are not protected at -98 bp. At no time was there any detectable protein DNA interaction in pFO003m5, m6, or m1 at Site II or the distal part of Site I. Lane 8, pFO003 plasmid DNA treated with DMS as a control for the G residue sequencing pattern. The only detectable protein binding occurs in Site I of pFO003m5 and m6 at the putative Sp1 site.


A.
E
Hc H
E
Site I
Site II
N
Hc
B.
1 2 3 4 5 6 7 8
-23
Site II
-63
-89
Site I
-123
-98

in the protein/DNA interactions in the heterologous mouse system. The results presented in Figure 4-6b, lane 3, suggested that in a cell line with low copy number, pFO003m5 (13 copies), a significant portion (approximately 70%) of the genes had protein bound to the proximal side of Site I (the putative Sp1 site G residues -98 to -100 protected in vivo), but there was apparently no protein bound to site II. When the copy number of the cell line increased to 20, pFO003m6, there was still protein/DNA interaction detectable at Site I, but not at Site II (lane 7). Finally when 140 copies of the human histone gene were present, pFO003m1, there was no detectable protein interaction at Site I or Site II (lane 7). As a control for the presence and location of Site I and Site II, synchronized HeLa cells, early in S phase, were treated with DMS at the same time and subjected to the same protocol as the pFO003 cell lines (lane 2). pFO003 plasmid DNA and deproteinized HeLa DNA were DMS treated as a control (lanes 8 and 1 respectively) for the expected sequence pattern of the G residues.

The results substantiated the cell line expression data that a limiting factor(s) was necessary for the transcription of histone genes. The results also support the contention that the mouse and human transcriptional proteins are not necessarily identical. Previous studies, including this work, have demonstrated that the mouse cell is able to correctly express introduced human histone genes (Green et al., 1986, Capasso and Heintz, 1985). In many other respects the mouse cell is capable of the regulation of human histone mRNA in a manner identical to that of the human cell. We have demonstrated that the processed 3′ ends of the human histone H4 mRNA are identical in mouse

and human cells and that the transcription initiation sites are also identical (data not shown). However, our failure to detect protein bound to the distal side of Site I and to all of Site II, even at low copy number, indicated that there were differences in the factors that bind there. Confirmation of their existence and binding in vitro has been demonstrated by van Wijnen et al. (1988, and personal communication). Perhaps there are subtle protein sequence variations that preclude detection with genomic sequencing. Previously, van Wijnen et al. (1987) demonstrated that there were factors that bound to the region of the H4 promoter from -210 to -410 bp, however these proteins were not detected in vivo by genomic sequencing. Either these protein/DNA complexes were artifacts of the in vitro assay system utilized or some protein/DNA interactions are simply not detectable with genomic sequencing. We should note that Dr. Pauli examined the region from -210 to -410 bp with both DNaseI and DMS protection. DNaseI is likely to detect a majority of the interactions, whereas DMS might not pick up every interaction (Dr. Pauli, personal communication).

We reanalyzed the copy number and expression data in light of the genomic sequencing results and found that in most, but not all cases, the human and mouse absolute densitometry signals inverted as the copy number increased. This is graphically presented in Figure 4-7 for several cell lines including pFO003. We calculated the percent of the total S1 nuclease protected fragment signal measured densitometrically (mouse H4 + human H4) that was representative of the mouse H4 gene and plotted this versus the copy number of each monoclonal cell line. It

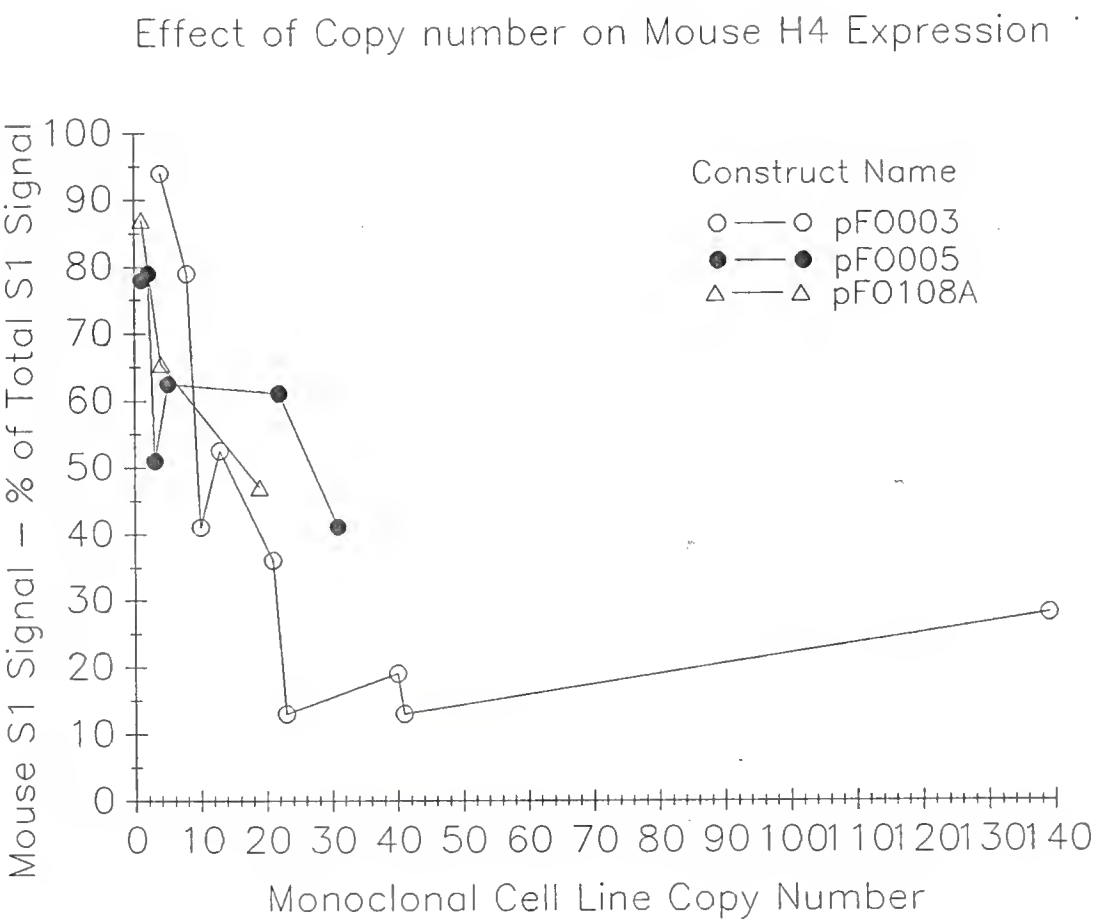


Figure 4-7 Effect of Human H4 gene copy number on Mouse H4 gene expression.

S1 nuclease protection data for several cell lines was analyzed to determine the effect of the human H4 gene on the expression of the mouse H4 gene. The human and mouse S1 nuclease assay densitometry values were totaled and the percent of the total signal that was mouse was plotted versus the copy number of the human H4 in each cell line. Data from pF0003, pF0005, and PF0108A are shown to illustrate the point that as human H4 copy number increases the expression of the mouse gene decreases.

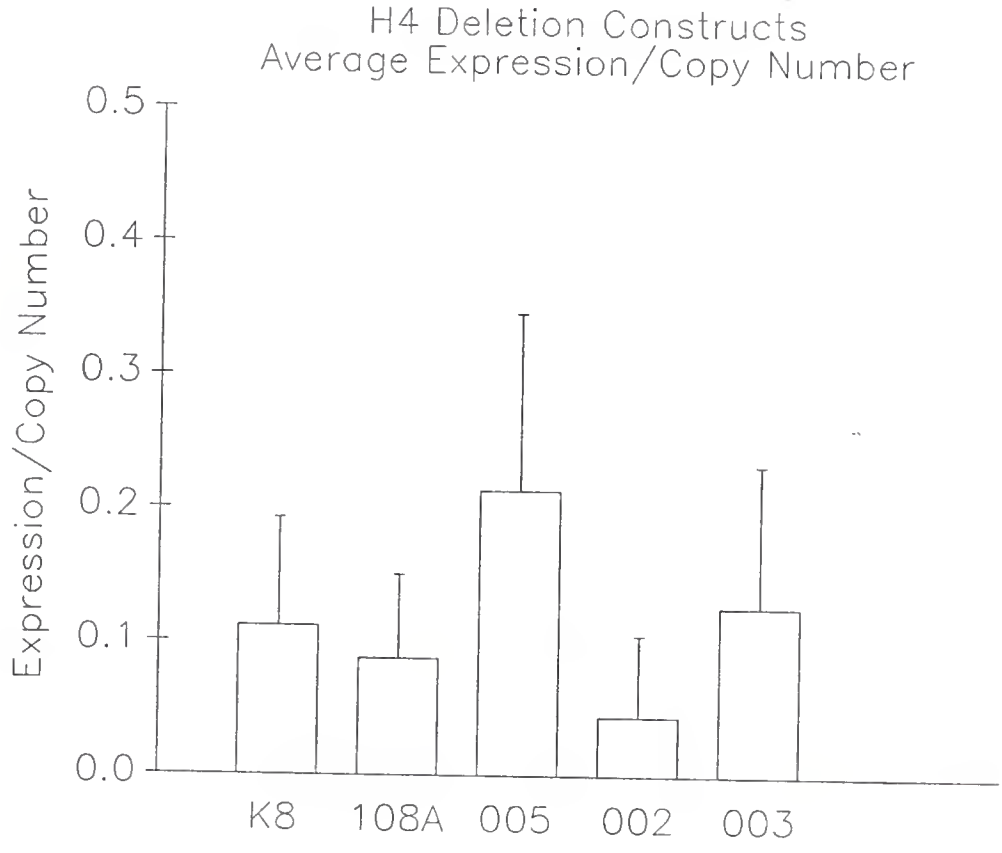


Figure 4-8 Reassessment of human H4 histone gene expression: low copy number data.

The same cell lines that were depicted in Figure 3-17 are shown here. Only data from low copy number cell lines has been included. This in general corresponds to less than 20 copies/cell. Expression/copy number is plotted with the standard deviation of the mean as a one way error bar.

was obvious that as the human H4 copy number increased the percent of total mouse S1 signal decreased proportionally. This result requires some qualifications. We had expected that as the human H4 copy number increased the expression would also increase. If there was no competition between the two sets of genes than the mouse signal should have been unaffected and remained stable since the mouse copy numberdoes not change. If that logic is followed an additional step than the mouse signal should have decreased with human copy number when measured as a ratio. However, the human H4 expression did not continue to rise with the human H4 copy number. Effectively, the increase in the human H4 gene copy number appears to have lowered the mouse gene expression and therefore artificially raised the human gene expression. The result of this phenomenon is that the original human H4/mouse H4 ratio that was calculated is certainly inaccurate in high copy number cell lines. We have noticed that in very high copy number cell lines, such as pFO004m1, both the human and mouse H4 genes are expressed at low levels (Figure 3-27). This is probably the result of factor distribution between the possible transcription units in such a manner that none of the genes has a full complement of proteins necessary for expression. Our reassessment of the expression data is presented Figure 4-8, and only incorporates data from each cell line in which the competition phenomenon (generally low copy number cell lines) was not readily apparent. The expression/copy was plotted as before in Figure 3-17. The statistical differences between constructs that were detailed earlier is still valid for this part of the data.

CHAPTER 5

DISCUSSION AND CONCLUSIONS

Our studies over the last several years have contributed to the general understanding of histone gene expression and of the expression of human genes in a heterologous system. The histone genes have been studied intensely for decades and only now are beginning to be understood. From the work we have presented here and the work done by others (Hanley et al., 1985; Sierra et al., 1983; Dailey et al., 1986; van Wijnen et al., 1987) it is clear that the histone H4 promoter is composed of several discrete DNA sequence elements, including the TATA box, CAAT box, Sp1 site (5'-GGCGGG-3'), and GGTCC element. We have also demonstrated that more distal sequences may have both a positive and negative effect on the transcriptional regulation of the FO108 human H4 histone gene.

We initially wanted to demonstrate what sequences were sufficient for cap site initiation of transcription <u>in vivo</u>. Previously, Sierra et al. (1983) had demonstrated with a series of Bal31 deletions, that sequences contained in the construct J67 (-47 bp), including the TATA box (-30 bp) and GGTCC element (-47 bp), were sufficient for correct initiation of transcription <u>in vitro</u>. In order to ascertain whether these sequences were sufficient <u>in vivo</u> we constructed a series of polyclonal cell lines in mouse C127 cells with the Bal 31 deletion constructs as described in Chapter 3. Transcription from each

construct was measured by S1 nuclease analysis and the expression level per copy determined. Six polyclonal cell lines of the construct J67 (-47 bp) were prepared and, although all cell lines contained detectable copies of the J67 construct as determined by Southern blot analysis, none of the cell lines initiated human H4 transcription correctly. An examination of the in vivo protein/DNA interactions within the proximal promoter region of the human H4 gene (Pauli et al., 1987) had previously revealed two sites of interaction, Site I (-124 bp to -89 bp) and Site II (-64 bp to -23 bp). We believe that the lack of correct in vivo transcription initiation from the J67 construct (-47 bp) is the result of the deletion of Site I and the distal half of Site II. Even though the GGTCC element (-47 bp) and the TATA box (-32 bp) are still present in the J67 construct, they are apparently insufficient for site specific transcription initiation in vivo. The GGTCC element that remains in the J67 construct is probably incapable of binding its respective protein. Pauli et al. (1987) have demonstrated that the in vivo factor interaction with the GGTCC element occurs symmetrically at three G residues on both DNA strands. The J67 deletion disrupts the symmetry of this binding through deletion of the distal G residue on the bottom strand. Additionally, the CAAT..2bp..GGTCC motif that is well conserved in many H4 histone genes (Wells, 1986) is disrupted by the J67 deletion, suggesting that it may also be important for transcriptional regulation. Our results suggest that multiple transcription factors are required for H4 transcription initiation and, in support of this hypothesis, van Wijnen et al. (1987, 1988) have demonstrated specific protein binding regions within Site I

and Site II of the FO108 H4 histone gene *in vitro*. We were able to demonstrate the necessity for all of the Site II protein/DNA interactions since correct initiation of transcription was observed with the construct J56 (-73 bp) that includes all of Site II. In contrast to the *in vitro* transcription results of Sierra et al. (1983), we have demonstrated that sequences between -47 and -73 bp, included in the construct J56 (-73 bp), are required for H4 histone transcription initiation *in vivo*.

The protein/DNA interactions at Site I (-124 bp to -89 bp) were shown by Pauli et al. (1987) to overlap a putative Sp1 binding site (5'-GGGGCGGGGC-3') as described by Briggs et al. (1986). We were interested to know whether this Sp1 site was functional and contributed to the transcriptional regulation of the FO108 human H4 histone gene. A cell line that contained the Bal 31 deletion construct J50 (-100 bp) was prepared and assayed by S1 nuclease analysis. With this cell line we demonstrated that the additional sequences between -73 and -100 bp, included in the construct J50, increased the level of *in vivo* transcription at least 10 fold above the construct J56 (-73 bp). Our result is consistent with interaction of Sp1 or an Sp1-like protein with this sequence and that this is responsible for the increase in transcription we have noted. Additionally, we have been able to demonstrate with genomic sequencing (Church and Gilbert, 1984) that there is a factor in mouse C127 cells that binds to the Sp1 recognition sequence *in vivo*, although we cannot conclude that it is indeed Sp1. Taken together, our results, and those of Pauli et al. (1987) and van

Wijnen et al. (1987, 1988) implicate Spl as a positive transcription factor in the regulation of this human H4 histone gene.

Because the 5′ flanking region of the FO108 H4 histone gene is very extensive, we characterized the contribution of more distal 5′ flanking sequences to the transcriptional regulation of this H4 histone gene *in vivo*. We first established that when all of Site I and Site II were present in cell lines that contained the construct K8 (-155 bp) no further increase in the level of transcription was detected. Extension of the promoter sequences to -215 bp in the construct pFO108A demonstrated that *in vivo* sequences from Site I (-124 bp) to -215 bp did not influence the level of transcription. These results were determined from experiments with both polyclonal and monoclonal cell lines of K8 and pFO108A.

The inclusion of sequences up to -417 bp in the construct pFO005 resulted in a 2-fold increase in the level of transcription above that demonstrated with the pFO108A construct cell lines. Previous analysis of this region by Pauli et al. (1987) had revealed no detectable *in vivo* protein/DNA interactions. In order to determine more precisely the location of the positive transcription element, two deletions of the pFO005 construct were prepared with Exonuclease III and assayed in monoclonal cell lines and short term transient expression experiments. The deletions have been denoted pFO005BSdel2-6 (-285 bp) and pFO005BSdel2-10 (-335 bp). Our results from the monoclonal cell lines constructed support the idea that the positive transcription element in pFO005 is located in the sequences from -215 bp (pFO108A) to -335 bp (pFO005BSdel 2-10). Comparison of the level of transcription from the

pF0005BS construct (-417 bp) and pF0005BSdel2-10 (-335 bp) demonstrated that the deletion (~ 80 bp) had not affected the level of transcription. Only a single monoclonal cell line was obtained with the construct pF0005BSdel2-6 (-285 bp) and the level of transcription was shown to be lower than pF0005BS. Because of the lack of appropriate cell lines we were unable to assess the effect of this deletion on the level of H4 histone gene transcription. Ken Wright of our laboratory has demonstrated similar transcription results in an in vitro transcription system with these constructs (personal communication). Preliminary analysis of the sequences from -215 bp to -335 bp suggested that secondary structure might be responsible for the function of this region. There are two possible inverted repeats within the region that may form stable stem and loop structures. Stable cell lines and short term transient expression experiments with the construct pF0001 (-3.3 kb, internal deletion -586 bp to -215 bp) also support the contention that the sequences from -417 bp to -215 bp (pF0005) contain a positive transcription element. In polyclonal and monoclonal cell lines the pF0001 construct is expressed at a significantly lower level than pF0005 and pF0108A. This result was duplicated in the transient expression experiments described in chapter 3.

We examined even more distal sequences and demonstrated that in stable monoclonal cell lines and short term transient expression assays the construct pF0007 (-586 bp) exhibits the same level of expression as pF0005 (-417 bp). If sequences extending to -1065 bp were included (pF0002), there was a significant (2-3 fold) decrease in the level of

transcription. Based on this observation we proposed that there was a negative regulatory element between -586 bp and -1065 bp in the histone H4 promoter. The region was sequenced by the method of Maxam and Gilbert (1980) in order help to identify any possible sequence elements that might be responsible for the decrease in transcription. Our analysis found two candidate sequences, Box 1 (-710 bp, 5'-TCCCCTCTCAG-3') and Box 2 (-580 bp, 5'-ATTCTCCTGT-3'), with homology to negative regulatory elements as described by Baniahmad et al. (1987) for the chicken lysozyme gene. To determine if these elements had any functionality we constructed two deletions in the 460 bp BamHI/EcoRI fragment of pFO002 designated pFO002D1 (-920 bp) and pFO002E9 (-730 bp). These constructs were assayed in comparison to pFO002, pFO007, pFO005, pFO001, and pFO108A for expression of the FO108 H4 histone gene in transiently transfected C127 and Ltk$^-$ mouse cells. In both cell types we demonstrated that the sequences we had proposed based on similarity were not responsible for the observed negative regulation. Both Box 1 and Box 2 were included in the construct pFO002E9, which was the most highly expressed construct of the group. Because pFO002E9 was expressed at a level approximately 2.5 fold higher than pFO007 we proposed that there was a positive element located between -586 bp and -730 bp. The only obvious candidate sequence present in the region was a CCAAT box located at -718 bp. We cannot conclude that this sequence is responsible for the increase in the transcriptional level of pFO002E9; however, it has been well documented that when located in the proximal promoter region of many genes the CCAAT box functions in the regulation of transcription in conjunction with other DNA sequences

(Dorn et al., 1987; McKnight and Kingsbury, 1982; McKnight et al. 1984; McKnight and Tjian, 1987).

Our experiments did indicate the existence of a negative regulatory element when we examined pFO002 (-1065 bp) and pFO002D1 (-920 bp). We demonstrated that these constructs were expressed at a significantly lower level in the short term transient assays than pFO002E9 (-730 bp). We therefore concluded that the negative regulatory element suggested by previous experiments more likely resided in the sequences between -730 bp and -920 bp. Dr. Chrysogelos, of our laboratory, identified a nuclease sensitive region (DNase I and S1) located between -720 bp to -820 bp that may represent a protein/DNA interaction (Dr. Susan Chrysogelos, personal communication). The sequence of this region contains a stretch from -800 bp to -960 bp was very A/T rich (70%).

We found that the region from -580 to -1010 bp contained two excellent homologies to MARs (matrix attachment regions) as described by Gasser and Laemmli (1987) and a topoisomerase II site (Sander and Hsieh, 1985). This topoisomerase II site was confirmed <u>in vitro</u> with purified enzyme by Dr. Tom Rowe (personal communication). Matrix attachment sites on the eukaryotic chromosome are thought to function in the regulation of gene expression through recognition of chromatin domains and attachment to the nuclear matrix as has been demonstrated for a number of genes, including Drosophila histone genes (Gasser and Laemmli, 1987). Since the histone genes of higher eukaryotes are clustered, it is possible that they may be divided into functional domains on chromatin loops. Cockerill et al. (1986) demonstrated that MARs were approximately 200 bp in length and 74% A/T. They also

demonstrated possible binding sites for topoisomerase II as described by Sander and Hsieh (1985). The upstream region of the FO108 gene from -800 to -960 is 70% A/T and contains at least one confirmed topoisomerase II site. This evidence suggests that binding to the nuclear matrix and DNA topology may function in the regulation of histone H4 gene expression.

Additionally, we demonstrated that pFO002 and pFO005 were expressed at nearly the same level in transiently transfected cells where the DNA was presumably episomal. However, in stable cell lines, pFO005 was expressed a significantly higher level (~ 3 fold) than pFO002. These results suggest that the state of the DNA, episomal or integrated, affects the function of certain DNA elements. This is consistent with our hypothesis that attachment to the nuclear matrix and DNA topology have a role in the regulation of the FO108 human H4 histone gene.

We have also noted that many of the negative regulatory elements previously described are located at a considerable distance from the gene they are associated with and this is consistent with our hypothesis (Baniahmad et al., 1987; Laimins et al., 1986). Additionally, experiments performed by Dr. Pauli, of our laboratory, suggested that histone H1, and a 43 kd nuclear acidic protein (non-histone) bound specifically to this region (Dr. Urs Pauli, personal communication). It has been previously suggested that histone H1 might be a general negative regulatory factor for transcription (Weintraub, 1985).

Another possibility that we have considered is that the strings of poly A_n and poly T_n may have an unusual secondary structure in the

upstream region of the H4 promoter. Although we have not been able to perform any direct analysis, it seems possible that under certain circumstances this segment of DNA might assume a "bent" conformation as described by Koo et al. (1986) and Travers (1987). It has been elegantly demonstrated in a number of systems, both prokaryotic and eukaryotic, that DNA can bend intrinsically if the necessary bases are present or can bend in response to the interaction of a protein (Salvo and Grindley, 1987; Koo et al., 1986). Bending of DNA requires that there be proper spacing between the AA dinucleotide pairs and poly A tracts. This spacing corresponds to approximately 10 bp, or a single turn of the helix (Koo et al., 1986). Several of the poly A_n tracts in the upstream region of the H4 gene from -945 bp to -880 bp appear appropriately spaced. This evidence suggests that the upstream region of the H4 promoter has unusual structure and might be responsible for the negative regulation we demonstrated.

We had previously implicated an additional positive regulatory element in the distal region of the H4 promoter. Preliminary polyclonal cell line experiments suggested that the pFO116 fragment (-6.0 to -7.5 kb) of λ HHG41 was able to enhance the level of transcription several fold. Helms et al. (1987) had shown that this fragment could stimulate CAT gene expression 4-5 times in HeLa cells when located at the 3' end of the gene. Their experiments suggested that the element might have the properties of enhancer. We examined this possibility further through the construction of a number of variant enhancer constructs. We made pFO004, pFO004R, and pFO006 as described in chapter 3 to test the hypothesis that this element had the

distance and orientation independent properties of an enhancer element. We examined a number of monoclonal cell lines prepared with each construct and established that the pFO116 fragment did not conclusively enhance the level of expression of the H4 gene in mouse C127 cells above that found with pFO108A or pFO108X. The construct pFO004 was highly expressed in a number of cell lines; however, this was apparently the result of high copy number and not enhancement of expression. The region was sequenced by Ken Wright and Dr. Urs Pauli and they found that the pFO116 fragment exhibited three sequences in the 500 bp EcoRI (-6.0 kb)/XbaI (-6.5 kb) section with strong similarity to the consensus core sequence of the SV40 and Ig heavy chain enhancers (Maniatis et al., 1987; Khoury and Gruss, 1983). We can only speculate that the lack of enhancer activity in mouse C127 cells, a fibroblast cell line, is due to the presence of negative regulatory factors or the absence of positive factors required for activity. Consistent with this idea, Wasylyk and Wasylyk (1986) demonstrated that the Ig heavy chain enhancer was negatively regulated in fibroblasts, but transcription could be stimulated in these cells if certain sequences were deleted. Finer analysis of the pFO116 fragment in different cell types should reveal if this element is regulated in such a manner.

Our studies of the pFO004 construct allowed us to associate repetitive sequences with higher copy number of the monoclonal cell lines and with specific integration. Most of our constructs integrated via the pathway described by Perucho et al. (1980). There was a cointegrate stage followed by integration at a limited number of sites

in the cellular chromatin. We found that minor repeated sequences, with some homology to the Alu repeat (Collart et al., 1985), located in the distal 5' flanking of pF0003 were responsible for the higher copy number in these cell lines. The pF0003 constructs had a considerable amount of tandem integration that also suggested that the minor repeats did not perturb the integration pathway. pF0004 was studied in more detail, and we concluded that the high copy number and heterogeneous integration observed was due to specific integration via the Alu repeat located in the pF0116 fragment (~ -7.0 kb). It has been previously demonstrated that the human histone genes are interspersed with various repeated sequences and often flanked by Alu repeats (Collart et al., 1985). Perhaps this unusual sequence organization accounts for the clustered but random organizational pattern of this family of genes.

When we examined the expression of the human H4 histone gene in the heterologous C127 cells we found that only a limited number of copies were expressed. In addition, we found that as copy number of the human H4 gene increased, the expression of the mouse H4 gene decreased. This observation has been made previously by Capasso and Heintz (1985) in which they found that in a cell line with a high copy number of the human H4 gene, pHuH4 (120 copies), the endogenous mouse H4 genes were completely shut off. Our results are similar and suggested competition for transcription factors. However, when the human H4 gene was present in very high copy number (cell lines pF0004ml, ~ 250 copies and pF0003ml, ~ 139 copies) neither the mouse nor human H4 genes were expressed to any significant extent. We feel that it is likely that

the regulatory molecules necessary for transcriptional control are limited, and when spread among a large number of transcription units, none of the units has a full complement. During the course of our studies we have demonstrated that the endogenous mouse H4 and transfected human H4 histone genes are in direct competition for a limited transcription factor. Genomic sequencing experiments described in chapter 4 demonstrated that we could detect binding in vivo of a protein to the Sp1 site (-100 bp) located in Site I. We were never able to demonstrate in vivo binding to Site II although the existence of the factors in mouse cells has been demonstrated in vitro by André van Wijnen (personal communication). The genomic sequencing experiments that we have described also demonstrated that the binding to the Sp1 site was titratable with increased copy number of the pFO003 construct. Binding was detected in cell lines pFO003M5 (~ 15 copies) and pFO003M6 (~ 25 copies) but not in pFO003M1 (~ 140 copies). This suggested that the interaction at the Sp1 site was titratable even though Sp1 is known to be an abundant transcription factor (Dynan and Tjian, 1985). Perhaps the binding of Sp1 to this sequence is dependent on the interaction with adjacent histone specific transcription factors that we have been unable to detect with genomic sequencing. Our inability to detect the protein/DNA ineractions at Site I in the mouse cell lines may simply reflect minor differences in analogous binding proteins between the mouse and human cells. These experiments support the contention that the H4 histone genes in the mouse cell lines directly competed for a limiting transcription factor or factors that function in the regulation of H4 histone gene expression in vivo.

In conclusion, our studies have described the functional role of transcription factor interactions seen at both Site I and Site II. Site II was required for initiation of transcription and Site I augmented the level of transcription in a positive manner. We concluded that in mouse C127 cells, an enhancer-like element in the far upstream region of the H4 promoter was not active. The possibility of a negative regulatory element was investigated and our results suggest that the sequences upstream of -730 bp are responsible. The sequences from -800 to -960 bp were shown to be 70% A/T and contain putative nuclear matrix attachment and topoisomerase II sites. The results of our studies suggest that the promoter of the F0108 H4 histone gene, as defined in vivo, may be more extensive than previously thought. Further deletion analysis and investigation will describe the specific sequences responsible for the transcriptional regulation of this gene in vivo.

APPENDIX A

SAMPLE COPY NUMBER CALCULATION

To determine the copy number of each monoclonal cell line the human H4 gene signal, mouse 18S ribosomal signal, and plasmid DNA standards were subject to densitometric analysis as described in Chapter II. Once completed, all the copy number blots were compared to each other visually and on each blot of equivalent length exposure an 18S ribosomal band was as the standard for that experiment. This decision involved comparison of many films and the photographs of the gels prior to transfer. Every effort was made to ensure that equivalent standards were picked from the different experiments.

An example calculation is given below for the cell line pFO108Am2. The densitometric values determined are listed below for each variable:

Ribosomal Standard (Rstd) = 6309 densitometry units (DU).
pFO108A Ribosomal value (SRstd) = 5891 DU.
pFO108A human value (HV) = 5138 DU.
10 pg (1.3 copies/diploid genome) control = 820 DU = 630 DU/copy.
50 pg (6.5 copies/diploid genome) control = 1923 DU = 295 DU/copy.
50 pg (13 copies/diploid genome) control = 3938 DU = 302 DU/copy.

$$\text{pFO108Am2 copy number} = \frac{\text{Rstd}}{\frac{\text{SRstd}}{\text{plasmid control}}} \times \text{HV} = \frac{6309}{\frac{5891}{295}} \times 5138 = 19$$

APPENDIX B

SAMPLE CALCULATION OF HUMAN H4 EXPRESSION

This example serves to illustrate how the mouse H4 internal control and the plasmid DNA markers were used to calculate the level of human H4 expression. The example presented here is for the cell line pFO005m6. Pertinent numbers are listed and then the calculation is done. Because of the differences in the intensities of S1 protected fragments the mouse and human H4 values had to be determined from different length exposures.

Human H4 densitometry value (17 day exposure) = 2840 DU.
Mouse H4 densitometry value (16 hour exposure) = 1397 DU.
pBR322/HpaII marker band #1 (16 hour exposure) = 1701 DU.
pBR322/HpaII marker band #1, 1:4 dilution (17 day exposure) = 5510 DU.

Calculation:

1) 5510 x 4 = 20040 units.
2) 20040 / 1701 = 12.95 (the fold difference from 16 hours to 17 days exposure).
3) 2840 / [1397 x 12.95] = 0.16 = human/mouse expression ratio.

APPENDIX C

TABLE OF CONSTRUCTS

Construct	5'sequence (bp)	Vector	Comments
J67	-47	pBR322	Bal 31 deletion of pF0108A
J56	-73	pBR322	Bal 31 deletion of pF0108A
J50	-100	pBR322	Bal 31 deletion of pF0108A
K8	-155	pBR322	Bal 31 deletion of pF0108A
L14	-185	pBR322	Bal 31 deletion of pF0108A
pF0108A	-215	pBR322	3' site = HindIII (+1877)
pF0108X	-215	pUC19	3' site = XbaI (+1107)
pF0005	-417	pUC13	3' site = PstI (+1677)
pF0007	-586	pUC19	3' site = XbaI
pF0002E9	-730	pUC19	3' site = XbaI
pF0002D1	-920	pUC19	3' site = PstI
pF0002	-1065	pUC8	3' site = PstI
pF0001	-215/-586 to -3300	pBR322	Internal deletion of pF0919
pF0003	-6500	pUC13	3' site = XbaI
pF0004	-215 + pF0116	pUC8	same orientation as genomic
pF0004R	-215 + pF0116	pUC8	opposite orientation
pF0006	-215 + 500 bp EcoRI/XbaI fragment from pF0116	pUC19	same orientation as genomic
pF0005BS	-417	Bluescript M13+	
pF0005BSdel2-10	-335	Bluescript M13+	Exonuclease III deletion
pF0005BSdel2-6	-285	Bluescript M13+	Exonuclease III deletion

REFERENCES

Artishevsky, A., Wooden. S., Sharma, A., Resendez, Jr., E., Lee, A. 1987. Cell-cycle regulatory sequences in a hamster histone promoter and their interactions with cellular factors. Nature. 328: 823-827.

Bacheler, L. T., Smith, K. D. 1976. Transcription of isolated mouse liver chromatin. Biochemistry. 15: 3281-3290.

Banerji, J., Olson, L, Schaffner, W. 1983. A lymphocyte enhancer is located downstream of the joining region in the immunoglobulin heavy chain genes. Cell. 33: 729-740.

Banerji, J., Rusconi, S., Schaffner, W. 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell. 27: 299-308.

Baniahmad, A., Muller, M., Steiner, Ch., Renkawitz, R. 1987. Activity of two different silencer elements of the chicken lysozyme gene can be compensated by enhancer elements. EMBO J. 6: 2297-2303.

Baumbach, L., Marashi, F., Plumb, M., Stein, G., Stein, J. 1984. Inhibition of DNA replication coordinately reduces cellular levels of core and Hi histone mRNAs: requirement for protein synthesis. Biochemistry. 23: 1618-1625.

Baumbach, L., Stein, G., Stein, J. 1987. Regulation of human histone gene expression: the contributions of transcriptional and post-transcriptional control to the coupling of histone mRNA stability with DNA replication. Biochemistry. 26: 6178-6187.

Bekhor, I., Kung, G. M., Bonner J. 1969. Sequence-specific interactions of DNA and chromosomal proteins. J. Mol. Biol. 39: 351-64.

Bendig, M. M., Hentschel, C. C. 1983. Transcription of sea urchin histone genes in HeLa cells. Nucl. Acids Res. 11: 2337-2346.

Bendig, M. M., Thomas, T., Folk, W. R. 1980. Regulatory mutants of polyoma virus defective in DNA replication and the synthesis of early proteins. Cell. 20: 401-409.

Benech, P., Mory, Y., Revel, M., Chebath, J. 1985. Structure of two forms of the interferon-induced (2'-5')oligo A synthetase of human cells based on cDNAs and gene sequences. EMBO J. 4: 2249-2256.

Benoist, C., Chambon, P. 1980. Deletions covering the putative promoter region of early mRNAs of simian virus 40 do not abolish T antigen expression. Proc. Natl. Acad. Sci. USA. 77: 3865-3869.

Berk, A. J., Sharp, P. A. 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digested hybrids. Cell. 12: 721-732.

Berk, A. J., Sharp, P. A. 1978. Spliced early mRNAS of simian virus 40. Proc. Natl. Acad. Sci. USA. 75: 1274-1278.

Bienz, M., Pelham, H. R. B. 1986. Heat shock regulatory elements function as an inducible enhancer in the xenopus hsp 70 gene and when linked to a heterologous promoter. Cell. 45: 753-760.

Bienz-Tadmor, B., Zakut-Houri, R., Libresco, S., Givol, D., Oren, M. 1985. The 5′ region of the p53 gene: evolutionary conservation and evidence for a negative regulatory element. EMBO J. 4: 3209-3213.

Birkenmeier, E. H., Brown, D. D., Jordan, E. 1978. A nuclear extract of Xenopus laevis oocytes that accurately transcribes 5S RNA genes. Cell. 15: 1077-1086.

Bloch, D. P., Macquigg, R. A., Brack, S. D., Wu, J-R. 1967. The synthesis of deoxyribonucleic acid and histone in the onion root meristem. J. Cell. Biol. 33: 451-466.

Borun, T. W., Scharff, M. D., Robbins, E. 1967. Rapidly labeled, polyribosome-associated RNA having the properties of histone messenger. Proc. Natl. Acad. Sci. USA. 58: 1977-1983.

Bourgeois, S., Cohn, M., Orgel, L. E. 1965. Suppression of and complementation among mutants of the regulatory gene of the lactose operon of Escherichia coli. J. Mol. Biol. 14: 300-302.

Breathnach, R., Chambon, P. 1981. Organization and expression of eukaryotic split genes coding for proteins. Ann. Rev. Biochem. 50: 349-383.

Briggs, M., Kadonaga, J., Bell, S., Tjian, R. 1986. Purification and biochemical characterization of the promoter-specific transcription factor, Sp1. Science. 234: 47-52.

Busslinger, M. Portmann, M., Irminger, J., Birnstiel, M. 1980. Ubiquitous and gene-specific regulatory 5′ sequences in sea urchin histone DNA clone coding for histone protein variants. Nucl. Acids Res. 8: 957-977.

Butler, W. B., Mueller, G. C. 1973. Control of histone synthesis in HeLa cells. Biochem. Biophys. Acta 294: 481-496.

Capasso, O., Heintz, N. 1985. Regulated expression of mammalian histone H4 genes in vivo requires a trans-acting transcription factor. Proc. Natl. Acad. Sci. USA. 82: 5622-5626.

Carozzi, N., Marashi, F., Plumb, M., Zimmerman, S., Zimmerman, A., Wells, J. R. E., Stein, G., Stein, J. 1984. Clustering of human H1 and core human histone genes. Science 224: 1115-1117.

Casey, J. L., Di Jeso, B., Rao, K. K., Rouault, T. A., Klausner, R. D., Harford, J. B. 1988. Deletional analysis of the promoter region of the human transferrin receptor gene. Nucl. Acids Res. 16: 629-646.

Cedar, H., Felsenfeld, G. 1973. Transcription of chromatin in vitro. J. Mol. Biol. 77:237-254.

Chambon, P. 1975. Eukaryotic nuclear RNA polymerases. Ann. Rev. Biochem. 44: 613-638.

Chrysogelos, S., Riley, D. E. Stein, G., Stein, J. 1985. A human histone H4 gene exhibits cell cycle dependent changes in chromatin structure correlating with its expression. Proc. Natl. Acad. Sci. USA. 82: 7535-7539.

Church, G., Ephrussi, A., Gilbert, W., Tonegawa, S. 1985. Cell-type-specific contacts to immunoglobulin enhancers in nuclei. Nature. 313: 798-801.

Church, G., Gilbert, W. 1984. Genomic Sequencing. Proc. Natl. Acad. Sci. USA. 81: 1991-1995.

Clark, S. J., Kreig, P. A., Wells, J. R. E. 1981. Isolation of a clone containing human histone genes. Nucl. Acids Res. 9: 1583-1590.

Cockerill, P. N., Garrard, W. T. 1986. Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. Cell. 44: 273-282.

Collart, D., Stein, G. S., Stein, J. 1985. A series of repetitive DNA sequences are associated with human core and H1 histone genes. Mol. Cell. Biochem. 67: 161-170.

Concino, M., Lee, R. F., Merryweather, J. P., and Weinmann, R. 1984. The adenovirus major late promoter TATA box and initiation site are both necessary for transcription in vitro. Nuc. Acids Res. 12:7423-7433.

Contreras, R. and Fiers, W. 1981. Initiation of transcription in permeable, SV40-infected or noninfected CV1 cells: evidence for multiple promoters of SV40 late transcription. Nuc. Acids Res. 9:215-236.

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C., Chambon, P. 1980. Promoter sequences of eukaryotic protein-coding genes. Science. 209: 1406-1414.

Dahmus, M. E., and Bonner J. 1965. Increased template activity of liver chromatin, a result of hydrocortisone administration. Proc. Natl. Acad. Sci. USA. 54: 1298-1302.

Dailey, L., Hanley, S. M., Roeder, R. G., Heintz, N. 1986. Distinct transcriptional factors bind specifically to two regions of the human histone H4 promoter. Proc. Natl. Acad. Sci. USA. 83: 7241-7245.

Davidson, I., Fromental, C., Augereau, P., Wildeman, A., Zenke, M., Chambon, P. 1986. Cell-type specific protein binding to the enhancer of simian virus 40 in nuclear extracts. Nature. 323: 544-548.

Demerac, M. 1956. A comparative study of certain loci in Salmonella. Cold Sprg. Harb. Symp. Quant. Biol. 21: 113-120.

Dorn, A., Bollekens, J., Staub, A., Benoist, C., Mathis, D. 1987. A multiplicity of CCAAT box-binding proteins. Cell. 50: 863-872.

Dynan, W. S., Sazer, S., Tjian, R. Schimke, R. T. 1986. The transcription factor Spl recognizes a DNA sequence in mouse dihydrofolate reductase promoter. Nature. 319: 246-248.

Dynan, W. S., Tjian, R., 1983a. Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. Cell. 32: 669-680.

Dynan, W. S., Tjian, R. 1983b. The promoter-specific transcription factor Spl binds to upstream sequences in the SV40 early promoter. Cell. 35: 79-87.

Dynan, W. S., Tjian, R. 1985. Control of eukaryotic messenger RNA synthesis by sequence-specific DNA binding proteins. Nature. 316: 774-778.

Efstratiadis, A., Posakony, J., Maniatis, T., Lawn, R., O'Connell, C., Spiritz, R., DeRiel, J., Forget, B., Weissman, S., Slightom, J., Blechl, A., Smithies, O., Baralle, R., Shoulders, C., Proudfoot, N. 1980. The structure and evolution of the human β-globin gene family. Cell. 21: 653-668.

Emmer, M., de Crombrugghe, B., Pastan, I., and Perlman, R. 1970. Cyclic AMP receptor protein of E. coli: Its role in the synthesis of inducible enzymes. Proc. Natl. Acad. Sci. USA. 66: 480-487.

Evans, T., DeChiara, T., Efstratiadis, A. 1988. A promoter of the rat insulin-like growth factor II gene consists of minimal control elements. J. Mol. Biol. 199: 61-81.

Farnham, P. J., Schimke, R. T. 1985. Transcriptional regulation of mouse dihydrofolate reductase in the cell cycle. J. Biol. Chem. 260: 7675-7680.

Feinberg, A., Vogelstein, B. 1983. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132: 6-13.

Fey, G., Lewis, J. B., Grodzicker, T. Bothwell, A. 1979. Characterization of a fused protein specified by the adenovirus 2-simian virus 40 hybrid Ad2+NDldp2. J. Virol. 30: 201-217.

Folger, K. R., Wong, E. A., Wahl, G., Capecchi, M. R. 1982. Patterns of integration of DNA microinjected in cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules. Mol. Cell. Biol. 2: 1372-1387.

Fowlkes, D. M., Shenk, T. 1980 Transcriptional control regions of the adenovirus VAI RNA gene. Cell. 22: 405-413.

Fromm, M., Berg, P. 1982. Deletion mapping of DNA regions required for SV40 early region promoter function in vivo. J. Mol. Appl. Genet. 1: 457-481.

Garel, A., Axel, R. 1976. Selective digestion of transcriptionally active ovalbumin genes from oviduct nuclei. Proc. Natl. Acad. Sci. USA. 73: 3966-70.

Gasser, S. M., Laemmli, U. K. 1987. A glimpse at chromosomal order. Trends Genet. 3: 16-22.

Gaub, M., Dierich, A., Astinotti, D., Touitou, I., Chambon, P. 1987. The chicken ovalbumin promoter is under negative control which is relieved by steroid hormones. EMBO J. 6: 2313-2320.

Gilbert, W., Müller-Hill, B. 1966. Isolation of the lac repressor. Proc. Natl. Acad. Sci. USA. 56: 1891-1898.

Gillies, S. D., Morrison, S. L., Oi, V. T., Tonegawa, S. 1983. A tissue specific enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell. 33: 717-728.

Gilmour, R. S., Paul, J. 1970. Role of non-histone components in determining organ specificity of rabbit chromatins. FEBS Lett. 9: 242-44.

Gilmour, R. S., Paul J. 1969. RNA transcribed from reconstituted nuclear protein is similar to natural RNA. J. Mol. Biol. 40: 137-39.

Gluzman, Y., Sambrook, J. F., Frisque, R. J. 1980. Expression of early genes of origin-defective mutants of SV40. Proc. Natl. Acad. Sci. USA. 77: 3898-3902.

Goodbourn, S., Burstein, H., Maniatis, T. 1986. The human β-interferon gene enhancer is under negative control. Cell. 45: 601-610.

Gorman, C. M., Moffat, L. F., Howard, B. H. 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. 2: 1044-1051

Gorman, C. M., Rigby, P. W., Lane, D. P. 1985. Negative regulation of viral enhancers in undifferentiated embryonic stem cells. Cell. 42: 519-526.

Gottschling, D. E., Cech, T. R., 1984. Chromatin structure of the molecular ends of Oxytricha macronuclear DNA: phased nucleosomes and a telomeric complex. Cell. 38: 501-510.

Graham, F., van der Eb, A. 1973. A new technique for the assay of infectivity of adenovirus 5 DNA. Virol. 52: 456-457.

Graves, B. J., Johnson, P. F., McKnight, S. L. 1986. Homologous recognition of a promoter domain common to the MSV LTR and the HSV tk gene. Cell. 44: 565-576.

Green, L., Van Antwerpen, R., Stein, J., Stein, G., Triputti, P., Emanuel, B., Selden, J., Croce, C. 1984. A major human histone gene cluster on the long arm of chromosome 1. Science. 226: 838-840.

Green, L., Schlaffer, I., Wright, K., Moreno, J. L., Berand, D., Hager, G., Stein, J., Stein, G. 1986. Cell cycle dependent expression of a stable episomal human histone gene is a mouse cell. Proc. Natl. Acad. Sci. USA. 83: 2315-2319.

Grosschedl, R., Birnstiel, M. 1980a. Spacer DNA sequences upstream of the TATAAATA sequence are essential for promotion of H2A histone gene transcription. Proc. Natl. Acad. Sci. USA. 7102-7106.

Grosschedl, R., Birnstiel, M. L. 1980b. Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants in vivo. Proc. Natl. Acad. Sci. USA. 77: 1432-1436.

Hanahan, D. 1983. Studies on transformation of Escherichia coli with plasmids. J. Mol. Biol. 166: 557-580.

Hanley, S. M., Bleecker, G. C., Heintz, N. 1985. Identification of promoter elements necessary for transcriptional regulation of a human histone H4 gene in vitro. Mol. Cell. Biol. 5: 380-389.

Harvey, R. P., Robins, A. J., and Wells, J. R. E. 1982. Independently evolving chicken histone H2b genes: identification of a ubiquitous H2b-specific 5' element. Nuc. Acids Res. 10:7851-7863.

Heintz, N., Zernik, M., Roeder, R. G. 1981. The structure of the human histone genes: clustered but not tandemly repeated. Cell. 24: 661-668.

Heintz, N., Sive, H. L., Roeder, R. G. 1983. Regulation of human histone gene expression: Kinetics of accumulation and changes in the rate of synthesis and in the half-lives of individual histone mRNAs during the HeLa cell cycle. Mol. Cell. Biol. 3: 539-550.

Heintz, N., Roeder, R. G. 1984. Transcription of human histone genes in extracts from synchronized HeLa cells. Proc. Natl. Acad. Sci. USA. 81: 1713-1717.

Helms, S. R., van Wijnen, A. J., Kroeger, P., Shiels, A., Stewart, C., Stein, J. L., Stein, G. S. 1987. Identification of an enhancer-like element upstream from a cell cycle dependent human H4 histone gene. J. Cell. Phys. 132: 552-558.

Hentschel, C., Birnstiel, M. L. 1981. The organization and expression of histone gene families. Cell. 225: 301-313.

Hope, I. A., Struhl, K. 1986. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell. 46: 885-894.

Hu, S-L., Manley, J. 1981. DNA sequence required for initiation of transcription in vitro from the major promotor of adenovirus-2. Proc. Natl. Acad. Sci. USA. 78: 820-824.

Huang, R. C. C., Bonner, J. 1962. Histone, a suppressor of chromosomal RNA synthesis Proc. Natl. Acad. Sci. USA. 48: 1216-22.

Isenberg, I. 1979. Histones. Ann. Rev. Biochem. 48: 159-191.

Ish-Horowicz, D., Burke, D. F. 1981. Rapid and efficient cosmid cloning. Nuc. Acids Res. 9: 2989-2998.

Jacob, F., Monod, J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. 3: 318-356.

Johnson, A. W., Hnilica, L. S. 1970. In vitro RNA synthesis and nuclear proteins of isolated sea urchin embryo nuclei. Biochim. Biophys. Acta. 224: 518-30.

Jones, K. A., Yamamoto, K. R., Tjian, R. 1985. Two distinct transcription factors bind to the HSV thymidine kinase promoter in vitro. Cell. 42: 559-572.

Jones, K., Kadonaga, J., Rosenfeld, P., Kelly, T., Tjian, R. 1987. A cellular DNA-binding protein that activates eukaryotic transcription and DNA replication. Cell. 48: 79-89.

Kadesh, T., Zervos, P., Ruezinsky, D. 1986. Functional analysis of the murine IgH enhancer: evidence for negative control of cell-type specificity. Nucl. Acids Res. 14: 8209-8221.

Kakidani, H., Ptashne, M. 1988. GAL4 activates gene expression in mammalian cells. Cell. 52: 161-167.

Kedinger, C., Gniazdowski, M., Mandel, Jr., J. L., Gissinger, F., Chambon, P. 1970. α-amantin: a specific inhibitor of one of two DNA-dependent RNA polymerase activities from calf thymus. Biochem. Biophys. Res. Comm. 38: 165-171.

Khoury, G., Martin, M. A., Lee, T. N. H., Danna, K. J. Nathans, D. 1973. A map of SV40 transcription sites expressed in productively infected cells, J. Mol. Biol. 78: 377-389.

Khoury, G., Gruss, P. 1983. Enhancer elements. Cell. 33:313-314.

King. P. V., Blakesley, R. W., 1986. Optimizing DNA ligations for transformation. Focus. 8:1-3.

Klein, S., Gerster, T., Picard, D., Radbruch, A., Schaffner, W. 1986. Evidence for transient requirement of the IgH enhancer. Nucl. Acids Res. 13: 8901-8912.

Kleinsmith, L., J. Stein, J., Stein, G. 1976. Dephosphorylation of nonhistone proteins specifically alters the pattern of gene transcription in reconstituted chromatin, Proc. Natl. Acad. Sci. USA. 73: 1174-78

Koo, H., Wu, H., Crothers, D. M. 1986. DNA bending at adenine-thymine tracts. Nature. 320: 501-506.

Kroeger, P., Stewart, C., Schaap, T., van Wijnen, A., Hirshman, J., Helms, S., Stein, G., Stein, J. 1987. Proximal and distal regulatory elements that influence in vivo expression of a cell cycle-dependent human H4 histone gene. Proc. Natl. Acad. Sci. USA. 84:3982-3986.

Laimins, L., Holmgren-Konig, M., Khoury, G. 1986. Transcriptional "silencer" element in rat repetitive sequences associated with the rat insulin 1 gene locus. Proc. Natl. Acad. Sci. USA. 83: 3151-3155.

Larsen, P. R., Harney, J. W., Moore, D. D. 1986. Repression mediates cell-type-specific expression of the rat growth hormone gene. Proc. Natl. Acad. Sci. USA. 83: 8283-8287.

Lichtler, A. C., Sierra, F., Clark, S., Wells, J. R. E., Stein, G. S., Stein J. L. 1982. The multiple H4 histone mRNAs of HeLa cells are encoded in different genes. Nature. 298: 195-198.

Lin, F-L., Sternberg, N. 1984. Homologous recombination between overlapping thymidine kinase gene fragments stably inserted into a mouse cell genome. Mol. Cell. Biol. 4: 853-861.

Lindberg, U., Persson, T., Philipson, L. 1972. Isolation and characterization of adenovirus mRNA in productive infection. J. Virol. 10: 909-919.

Linxweiler, W., Horz, W. 1985. Reconstitution experiments show that sequence specific histone-DNA interactions are the basis for the nucleosome phasing on mouse satellite DNA. Cell. 42:281-290.

Loyter, A., Scangos, G. A., Ruddle, F. H. 1982. Mechanisms of DNA uptake by mammalian cells: Fate of exogenously added DNA monitored by the use of fluorescent dyes. Proc. Natl. Acad. Sci. USA. 79: 422-426.

Mandel, J. L., Chambon, P. 1970. Animal DNA-dependent RNA polymerases. Analysis of the RNAs synthesized on Simian Virus 40 supercoiled DNA by mammalian RNA polymerases AI and B. Eur. J. Biochem. 41: 379-95.

Maniatis, R. Fritsch, E. F., Sambrook, J. 1982. In Molecular cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Maniatis, T., Goodbourn, S., Fischer, J. 1987. Regulation of inducible and tissue-specific gene expression. Science. 236: 1237-1245.

Manley, J. L., Fire, A., Cano, A., Sharp, P. A., Gefter, M. L. 1980. DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Natl. Acad. Aci. USA. 77: 3855-3859.

Marashi, F., Helms, S., Shiels, A., Silverstein, S., Greenspan, D., Stein, G., Stein, J. 1986. Enhancer-facilitated expression of prokaryotic and eukaryotic genes using human histone gene 5' regulatory sequences. Biochem. Cell. Biol. 64: 277-289.

Maxam, A. M., Gilbert, W. 1980. Sequencing end-labelled DNA with base specific chemical cleavages. Meth. Enzymol. 65: 499-560.

Maxson, R., Cohn, R., Kedes, L., Mohun, T. 1983. Expression and organization of histone genes. Annu. Rev. Biochem. 17: 239-277.

McKnight, S. L., Kingsbury, R. 1982. Transcriptional control signals of a eukaryotic protein-coding gene. Science. 217: 316-324.

McKnight, S., Tjian, R. 1986. Transcriptional selectivity of viral genes in mammalian cells. Cell. 46: 795-805.

McKnight, S. L., Kingsbury, R. C., Spence, A., Smith, M. 1984. The distal transcription signals of the herpesvirus tk gene share a common hexanucleotide control sequence. Cell. 37: 253-262.

Melton, D. W., McEwan, C., McKie, A. B., Reid, A. M. 1986. Expression of the mouse HPRT gene: deletional analysis of the promoter region of an X-chromosome linked house-keeping gene. Cell. 44: 319-328.

Mitchell, P. J., Wang, C., Tjian, R. 1987. Positive and negative regulation of transcription in vitro: Enhancer-binding protein AP-2 is inhibited by SV40 T antigen. Cell. 50: 847-861.

Modrich, P., Anraku, Y., Lehman, I. R. 1973. Deoxyribonucleic acid ligase: Isolation and physical characterization of the homogeneous enzyme from Escherichia coli. J. Biol. Chem. 248: 7495-7501.

Moorman, A. F. M., deLaaf, R. T. M., Destree, O. H. J., Telford, J., Birnstiel, M. L. 1980. Histone genes from Xenopus laevis: molecular cloning and initial characterization. Gene. 10: 185-193.

Moreno, J. L., Chrysogelos, S. A., Stein, G. S., Stein, J. L. 1986. Reversible changes in the nucleosomal organization of a human H4 histone gene during the cell cycle. Biochemistry. 25: 5364-5370.

Muglia, L., Rothman-Denes, L. 1986. Cell type-specific negative regulatory element in the control region of the rat alpha-fetoprotein gene. Proc. Natl. Acad. Sci. USA. 83: 7653-7657.

Myers, R. M., Tilly, K., Maniatis, T. 1986. Fine structure genetic analysis of a β-globin promoter. Science. 232: 613-618.

Nasmyth, K. 1985. A repetitive DNA sequence that confers cell-cycle START (CDC-28)-dependent transcription of the HO gene in yeast. Cell. 42: 225-235.

Nathans, D., Smith, H. 1975. Restriction endonucleases in the analysis and restructuring of DNA molecules. Ann. Rev. Biochem. 44: 273-293.

Ng, S. Y., Parker, C. S., Roeder, R. G. 1979. Transcription of cloned Xenopus 5S RNA genes by X. laevis RNA polymerase III in reconstituted systems. Proc. Natl. Acad. Sci. USA. 76: 136-140.

Nir., U., Walker, M. D., Rutter, W. J. 1986. Regulation of rat insulin 1 gene expression: evidence for negative regulation in nonpancreatic cells. Proc. Natl. Acad. Sci. USA. 83: 3180-3184.

Park, W. D., Stein, J. L., Stein, G. S. 1976. Activation of in vitro histone gene transcription from HeLa S3 chromatin by s-phase nonhistone chromosomal proteins. Biochemistry 15: 3297-3300.

Paul, J., Gilmour, R. S. 1966. Template activity of DNA is restricted in chromatin. J. Mol. Biol. 16: 242-44.

Paul, J., Gilmour, R. S. 1968. Organ-specific restriction of transcription in mammalian chromatin. J. Mol. Biol. 34: 305-16.

Pauli, U., Chrysogelos, S., Stein, G., Stein, J., Nick, H. 1987. Protein-DNA interactions in vivo upstream of a cell cycle-regulated human H4 histone gene. Science. 236: 1308-1311.

Perucho, M., Hanahan, D., Wigler, M. 1980. Genetic and physical linkage of exogenous sequences in transformed cells. Cell 22: 309-317.

Perucho, M., Wigler, M. 1981. Linkage and expression of foreign DNA in cultured animal cells. Cold Spring Harbor Symp. Quant. Biol. 45: 829-838.

Picard, D., Schaffner, W. 1984. A lymphocyte-specific enhancer in the mouse immunoglobulin kappa gene. Nature. 307: 80-82.

Plumb, M. A., Stein, J. L., Stein, G. 1983a. Influence of DNA synthesis inhibition on the coordinate expression of core human histone genes during S phase. Nucl. Acids Res. 11: 7927-7945.

Plumb, M., Stein, J., Stein, G. 1983b. Coordinate regulation of multiple histone mRNAs during the cell cycle in HeLa cells. Nucl. Acids Res. 11: 2391-2410.

Prescott, D. M. 1966. The synthesis of total macronuclear protein, histone and DNA during the cell cycle in Euplotes erystomus. J. Cell. Bio. 31: 1-9.

Pribnow, D. 1975. Bacteriophage T7 early promoters: nucleotide sequences of two RNA polymerase binding sites. J. Mol. Biol. 99: 419-443.

Ptashne, M. 1986. Gene regulation by proteins acting nearby and at a distance. Nature. 322: 697-701.

Pulm, W., Knippers, R. 1985. Transfection of mouse fibroblast cells with a promoterless Herpes Simplex Virus thymidine kinase gene: number of integrated gene copies and structure of single and amplified gene sequences. Mol. Cell. Biol. 5: 295-304.

Remmers, E. F., Yang, J. Q., Marcu, K. B. 1986. A negative transcriptional control element located upstream of the murine c-myc gene. EMBO J. 5: 899-904.

Reynolds, G. A., Baser, S. K., Osborne, T. F., Chin, D. J., Gill, G., Brown, M. S., Goldstein, J. L., Luskey, K. L. 1984. HMG coA Reductase: a negatively regulated gene with unusual promoter and 5' untranslated regions. Cell. 38: 275-285.

Rigby, P. W. J., Dieckman, M., Rhodes, C., Berg, P. 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. 113: 237-251.

Roeder, R. G. 1976. Eukaryotic nuclear RNA polymerase. In: Losick, R., Chamberlin (eds) RNA Polymerase, Cold Spring Harbor, New York.

Rosales, R., Vigneron, M., Macchi, M., Davidson, I., Xiao, J. H., Chambon, P. 1987. In vitro binding of cell-specific and ubiquitous nuclear proteins to the octamer motif of the SV40 enhancer and related motifs present in other promoters and enhancers. EMBO J. 6: 3015-3025.

Rosenberg, M., Court, D. 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Ann. Rev. Genetics. 13: 319-353.

Ross, J., Peltz, S., Kobs, G., Brewer, G. 1986. Histone mRNA degradation in vivo: the first detectable step occurs at or near the 3' terminus. Mol. Cell. Biol. 6: 4362-4371.

Ross, J., Kobs, G., Brewer, G., Peltz, S. 1987. Properties of the exonuclease activity that degrades H4 histone mRNA. J. Biol. Chem. 262: 9374-9381.

Rowe, W. P., Huebner, R. J., Gillmore, L. K., Parrott, R. H., Ward, T. G. 1953. Isolation of a cytogenic agent from human adenoids undergoing spontaneous degeneration in tissue culture. Pros. Soc. Exp. Biol. Med. 84: 570-573.

Salvo, J., Grindley, N. 1987. Helical phasing between DNA bends and the determination of bend direction. Nucl. Acids Res. 15: 9771-9779.

Sambrook, J., Sugden, B., Keller, W., Sharp, P. A. 1973. Transcription of simian virus 40. III. Mapping of "early" and "late" species of RNA. Proc. Natl. Acad. Sci. USA. 70: 3711-3715.

Sander, M., Hsieh, T. S., 1985. Drosophila topoisomerase II double-strand DNA cleavage: analysis of DNA sequence homology at the cleavage site. Nuc. Acids Res. 13: 1057-1072.

Sawadago, M. S., Roeder, R. G. 1985. Factors involved in specific transcription by human polymerase II: analysis by a rapid and quantitative in vitro assay. Proc. Natl. Acad. Sci. USA. 82: 4394-4398.

Seiler-Tuyns, A., Birnstiel, M. L. 1981. Structure and expression in L-cells of a cloned H4 histone gene of the mouse. J. Mol. Biol. 151: 607-625.

Sen, R., Baltimore, D. 1986. Multiple nuclear factors interact with the immunoglobulin enhancer sequences. Cell. 46: 705-716.

Serfling, E. Jasin, M., Schaffner, W. 1985. Enhancers and eukaryotic gene transcription. Trends Genet. 1: 224-230.

Sharp, P. A., Gallimore, P. H., Flint, S. J. 1975. Mapping of adenovirus 2 RNA sequences in lytically infected cells and transformed cell lines. Cold Spring Harbor Symp. Quant. Biol. 39: 457-474.

Shenk, T. 1981. Transcriptional control regions: nucleotide sequence requirements for initiation by RNA polymerase II and III. Current Topics in Microbiology and Immunology. 93: 25-46.

Sierra, F., Lichter, A., Marashi, F., Rickles, R., Van Dyke, T., Clark, S., Wells, J., Stein, G., Stein J. 1982. Organization of human histone genes. Proc. Natl. Acad. Sci. USA. 79: 1795-1799.

Sierra, F., Stein, G., Stein, J. 1983. Structure and in vitro transcription of a human H4 histone gene. Nucl. Acids. Res. 11 7069-7086.

Simpson, R. J. 1973. Structure and function of chromatin. Adv. Enzymol. 38: 41-108.

Singer-Sam, J., Keith, D. H. Tani, K., Simmer, R. L., Shively, L., Lindsay, S., Yoshida, A., Riggs, A. D. 1984. Sequence of the promoter of the gene for human X-linked 3-phosphoglycerate kinase. Gene. 32: 409-417.

Sittman, D. B., Graves, R. A., Marzluff, W. F. 1983. Histone mRNA concentrations are regulated at the level of transcription and mRNA degradation. Proc. Natl. Acad. Sci. USA. 80: 1849-1853.

Sive, H. L., Heintz, N., Roeder, R. G. 1986. Multiple sequence elements are required for maximal in vitro transcription of a human histone H2B gene. Mol. Cell. Biol. 6: 3329-3340.

Smith, H. O., Wilcox, K. W. 1970. A restriction enzyme from Hemophilus influenzae. I. Purification and general properties. J. Mol. Biol. 51: 379-391.

Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gen electrophoresis. J. Mol. Biol. 98: 503-517.

Southern, P. J., Berg, P. 1982. Transformation of animal cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. J. Mol. Appl. Genet. 1: 327-341.

Stein, G. S., Spelsberg, T. C., Kleinsmith, L. J. 1974. Nonhistone chromosomal proteins and gene regulation. Science 183: 817-24.

Stein, G., Park, W., Thrall, C., Mans, R., Stein, J. 1975. Regulation of cell cycle stage-specific transcription of histone genes from chromatin by nonhistone chromosomal proteins. Nature 257: 764-67.

Stein, G. S., Sierra, F., Stein, J. L., Plumb, M. A., Marashi, F., Carozzi, N., Prokopp, K., Baumbach, L. 1984. Organization and expression of human histone genes. In Histone Genes pp. 397-455, John Wiley and Sons, New York.

Stein, G., Stein, J. 1984. Is human histone gene expression autogenously regulated? Mol. and Cell. Biochem. 64: 105-110.

Stumph, W. E., Hodgson, C. P., Tsai. M. J., O'Malley, B. W. 1984. Genomic structure and possible retroviral origin of the chicken CR1 repetitive DNA sequence family. Proc. Natl. Acad. Sci. USA. 81: 6667-6671.

Sugarman, B. J., Dodgson, J. and Engel, J. D. 1983. Genomic organization, DNA sequence and expression of chicken embryonic histone genes. J. Biol. Chem. 258: 9005-9016.

Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., Stahl, F. W. 1983. The double-strand-break repair model for recombination. Cell. 33: 25-35.

Tartof, K. D., Hobbs, C. A. 1987. Improved media for growing plasmid and cosmid clones. Focus 9:12.

Theisen, M., Stief, A., Sippel, A. E. 1986. The lysozyme enhancer: cell-specific activation of the chicken lysozyme gene by a far-upstream DNA element. EMBO J. 5: 719-724.

Thomas, M., White, R. L., and Davis, R. W. 1976. Hybridization of RNA to double-stranded DNA: Formation of R-loops. Proc. Natl. Acad. Sci. USA. 73: 2294-2298.

Thomas, D. C., and Green, M. 1966. Biochemical studies on adenovirus multiplication. XI. Evidence for a cytoplasmic site for the synthesis of viral coded proteins. Proc. Natl. Acad. Sci. USA. 56: 243-246.

Tooze, J. 1980. Molecular Biology of Tumor Viruses: DNA Tumor Viruses (2nd Ed. part 2 revised). Cold Sprg. Harb. Cold Spring Harbor Laboratory.

Travers, A. A. 1987. DNA bending and nucleosome positioning. Trends Biochem. Sci. 12: 108-113.

Treisman, R., Maniatis, T. 1985. Simian virus 40 enhancer increases the number of RNA polymerase II molecules on linked DNA. Nature. 315: 72-75.

Triputti, P., Emanuel, B., Croce, C., Green, L., Stein, G., Stein J. 1986. Human histone genes map to multiple chromosomes Proc. Natl. Acad. Sci. USA. 83: 3185-3188.

Tsai, M. J., Towle, H. C., Harris, S. E., O'Malley, B. W. 1976. Effect of estrogen on gene expression in the chick oviduct. Comparative aspects of RNA chain initiation in chromatin using homologous versus Escherichia coli RNA polymerase. J. Biol. Chem. 251: 1960-68.

Tullis, R. H., Rubin, H. 1980. Calcium protects DNaseI from proteinase K: A new method for the removal of contamnating RNase from DNAseI. Anal. Biochem. 107: 260-264.

van Wijnen, A. J., Stein, G. S., Stein J. L. 1987. A nuclear protein with affinity for the 5' flanking region of a cell cycle dependent human histone H4 gene in vitro. Nucl. Acids Res. 15: 1679-1698.

van Wijnen, A. J., Wright, K. L., Massung, R. F., Gerretsen, J., Stein, J. L., Stein, G. S. 1988. Two target sites for protein binding in the promoter region of a cell cycle regulated human H1 histone gene. Nucl. Acids Res. 16: 571-592.

Wasylyk B., Derbyshire, R., Guy, A., Molko, D., Roget, A., Teolue, R., Chambon, P. 1980. Specific in vitro transcription of conalbumin gene is drastically decreased by a single-point mutation in TATA box homology sequence. Proc. Natl. Acad. Sci. USA. 77: 7024-7028.

Wasylyk, C., Wasylyk, B. 1986. The immunoglobulin heavy-chain B lymphocyte enhancer efficiently stimulates transcription in non-lymphoid cells. EMBO J. 5: 553-560.

Webster, N., Jin, J. R., Green, S., Hollis, M., Chambon, P. 1988. The yeast UAS_G is a transcriptional enhancer in human hela cells in the presence of the GAL4 trans-activator. Cell. 52: 169-178.

Weil, P. A., Segall, J., Harris, B., Ng S-Y, Roeder, R. G. 1979. Faithful transcription of eukaryotic genes by RNA polymerase III in systems reconstituted with purified DNA templates. J. Biol. Chem. 254: 6163-6173.

Weintraub, H., Groudine, M. 1976. Chromosomal subunits in active genes have an altered conformation. Globin genes are digested by deoxyribonuclease I in red blood cell nuclei but not in fibroblast nuclei. Science. 193: 848-56.

Weintraub, H. 1985. Assembly and propagation of repressed and derepressed chromosomal states. Cell. 42: 705-711.

Weiss, B., Richardson, C. C. 1967. Enzymatic breakage and joining of deoxyribonucleic acid. I. Repair of single-strand breaks in DNA by an enzyme system from E. coli infected with T4 bacteriophage. Proc. Natl. Acad. Sci. USA. 57: 1021-1028.

Wells, D., Kedes, L. 1985. Structure of a human histone cDNA: evidence that basally expressed histone genes have intervening sequences and encode polyadenylated mRNAs. Proc. Natl. Acad. Sci. USA. 82: 2834-2838.

Wells, D. E. 1986. Compilation analysis of histones and histone genes. Nucl Acids Res. 14: 119-149.

Wu, G. J. 1978. Adenovirus DNA-directed transcription of 5.5S RNA in vitro. Proc. Natl. Acad. Sci. USA. 75: 2175-2179.

Wu, R. S., Bonner, W. M. 1982. Patterns of histone variant synthesis can distinguish Go from Gl cells. Cell. 31: 367-374.

Yee., A. S., Reichel, R., Kovesdi, I., Nevins, J. R. 1987. Promoter interaction of the E1A-inducible factor E2F and its potential role in the formation of a multi-component complex. EMBO J. 6: 2061-2068.

Zhong, R., Roeder, R. G., Heintz, N. 1983. The primary structure and expression of four cloned human histone genes. Nucleic Acids Res. 11: 7409-7425.

BIOGRAPHICAL SKETCH

Paul Edmond Kroeger was born in Louisville KY, on January 7, 1960. He was able to see much of the United States before the age of 13 when his family finally settled in Winston-Salem, NC. He attended R. J. Reynolds Senior High School where he graduated in the spring of 1978. He entered Wake Forest University in the fall of 1978 and graduated with a B.A. in biology in 1982. In the fall of 1982 he entered the graduate program in the Department of Immunology and Medical Microbiology and was partially supported in his studies by a training grant from the National Institutes of Health. He spent 2 and a half years in the lab of Dr. William Holloman studying eukaryotic recombination, and the remainder of his time in the lab of Janet and Gary Stein pursuing his thesis project. In the fall of 1987 he married Carol Ward and they produced a gorgeous son, Alan Scott Kroeger, the following spring. After graduation he will pursue a postdoctoral fellowship with Dr. Thomas Rowe in the Department of Pharmacology at the University of Florida.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Janet L. Stein, Chair
Professor of Immunology and
Medical Microbiology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Gary S. Stein
Professor of Biochemistry
and Molecular Biology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Richard W. Moyer
Professor Immunology and
Medical Microbiology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

William W. Hauswirth
Professor of Immunology and
Medical Microbiology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Harry Ostrer
Assistant Professor of
Biochemistry and Molecular
Biology

This dissertation was submitted to the Graduate Faculty of the College of Medicine and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August 1988

Dean, College of Medicine

Dean, Graduate School